



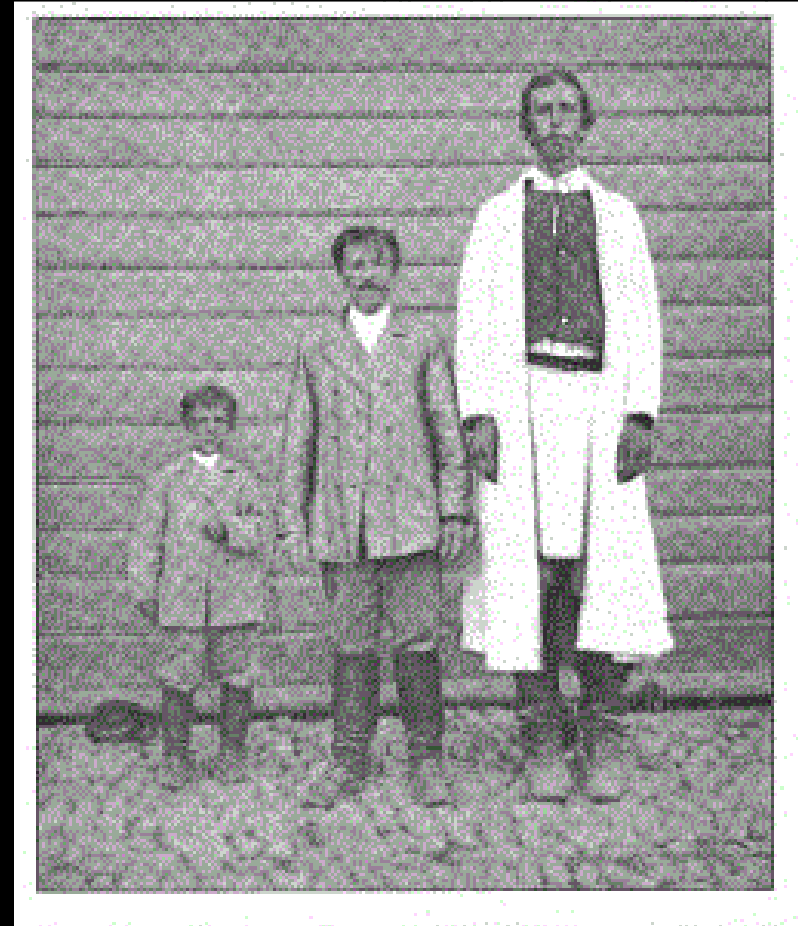
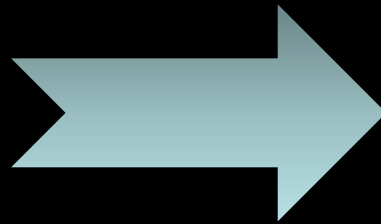
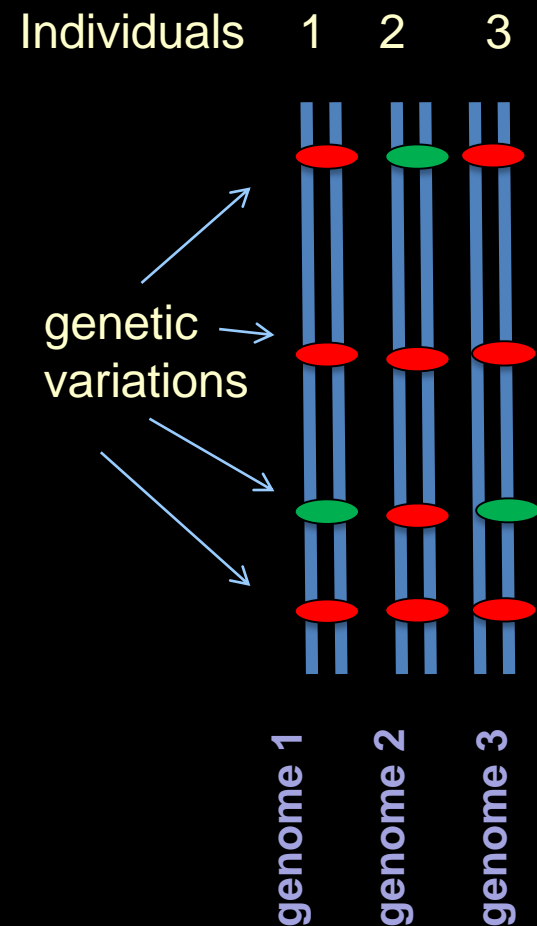
# Genetics variations – effects, propagation, and buffering

Teresa Przytycka  
NIH / NLM / NCBI



# Phenotypes

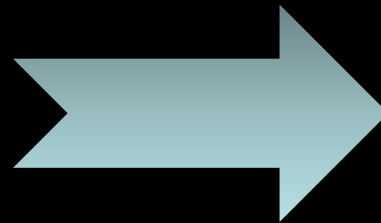
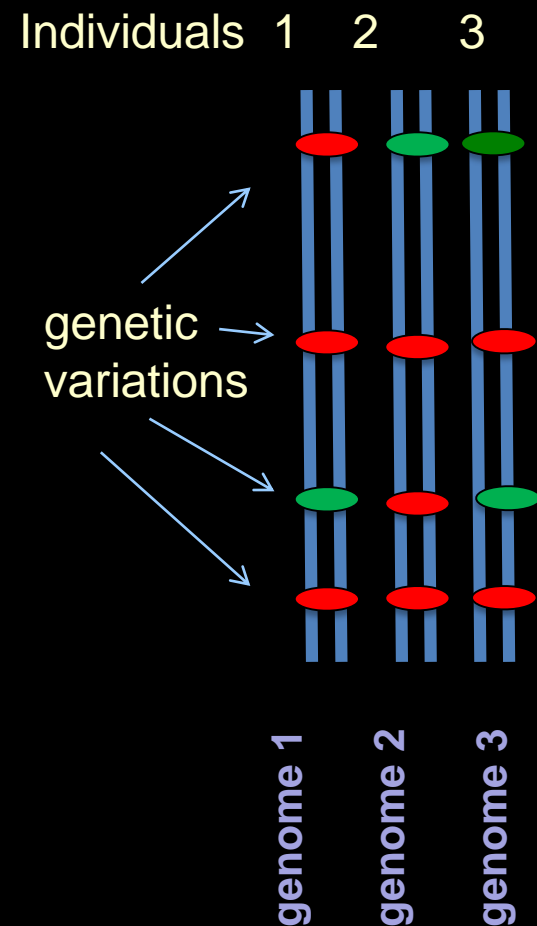
## Genotypes



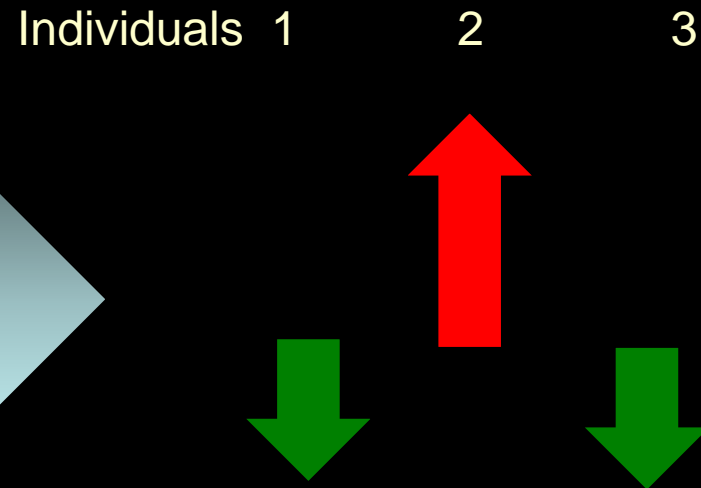
*Journal "Wisla" (1902)*

# Molecular Phenotypes

## Genotypes



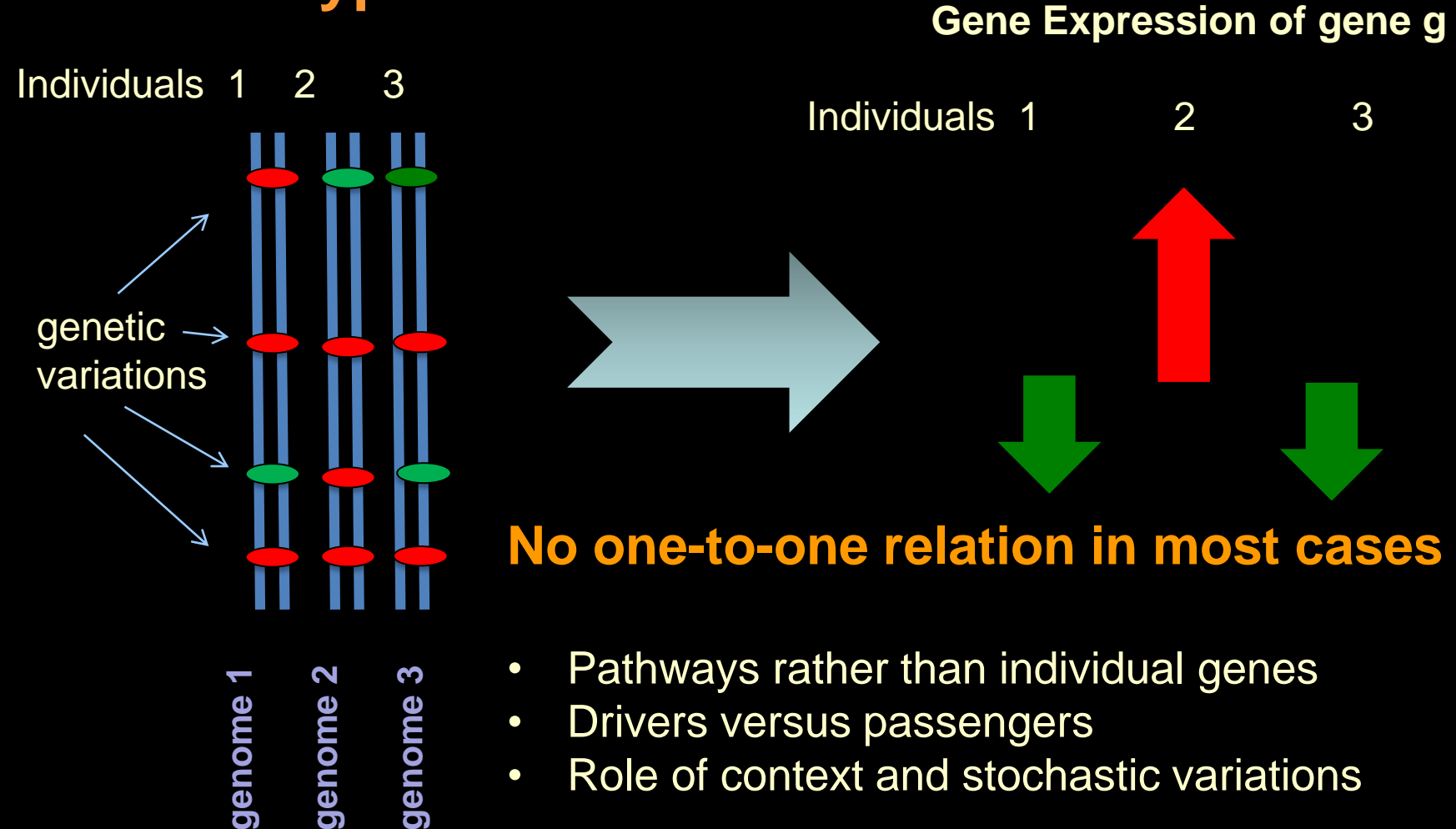
## Gene Expression of gene g



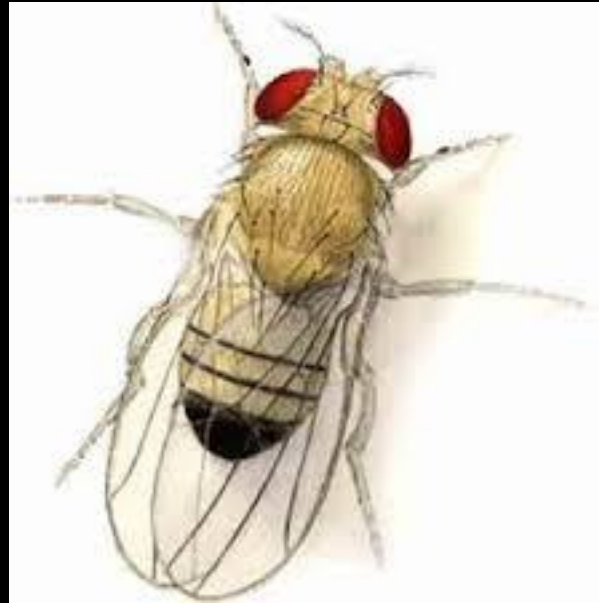
eQTL analysis

# Molecular Phenotypes

## Genotypes

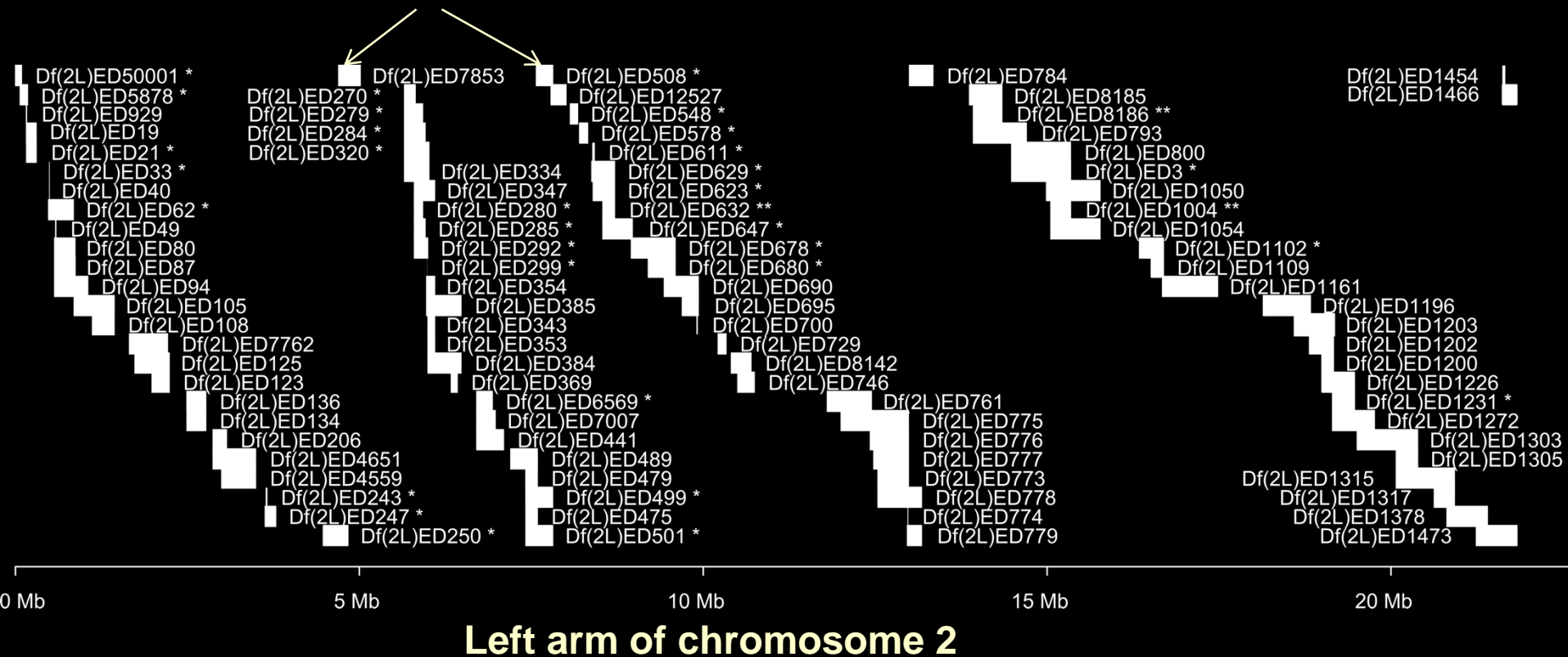


# Lessons from engineered genetic perturbations in Fly



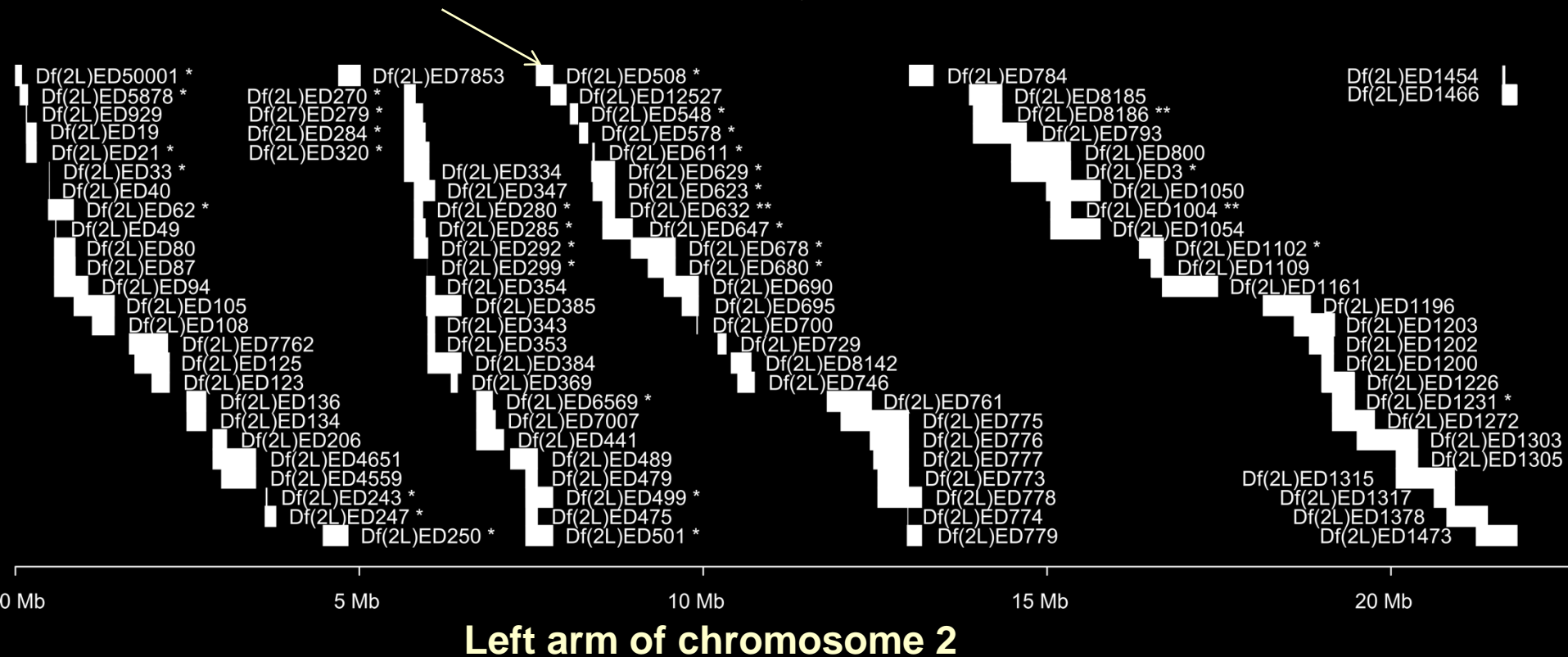
# Engineered chromosomal deletions

Chromosomal deletions (in one chromosome only)



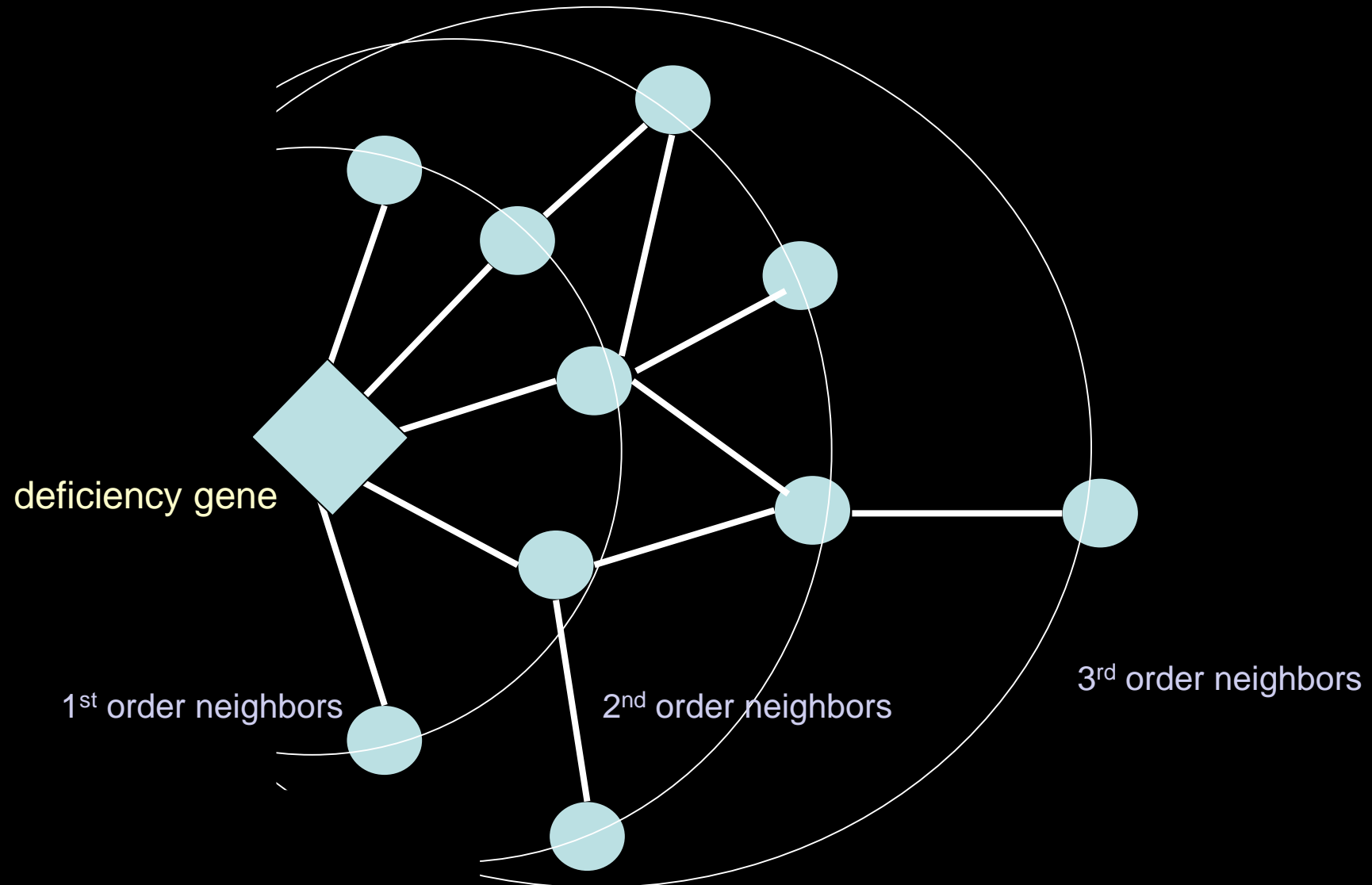
# Engineered chromosomal deletions

Chromosomal deletions (in one chromosome only)



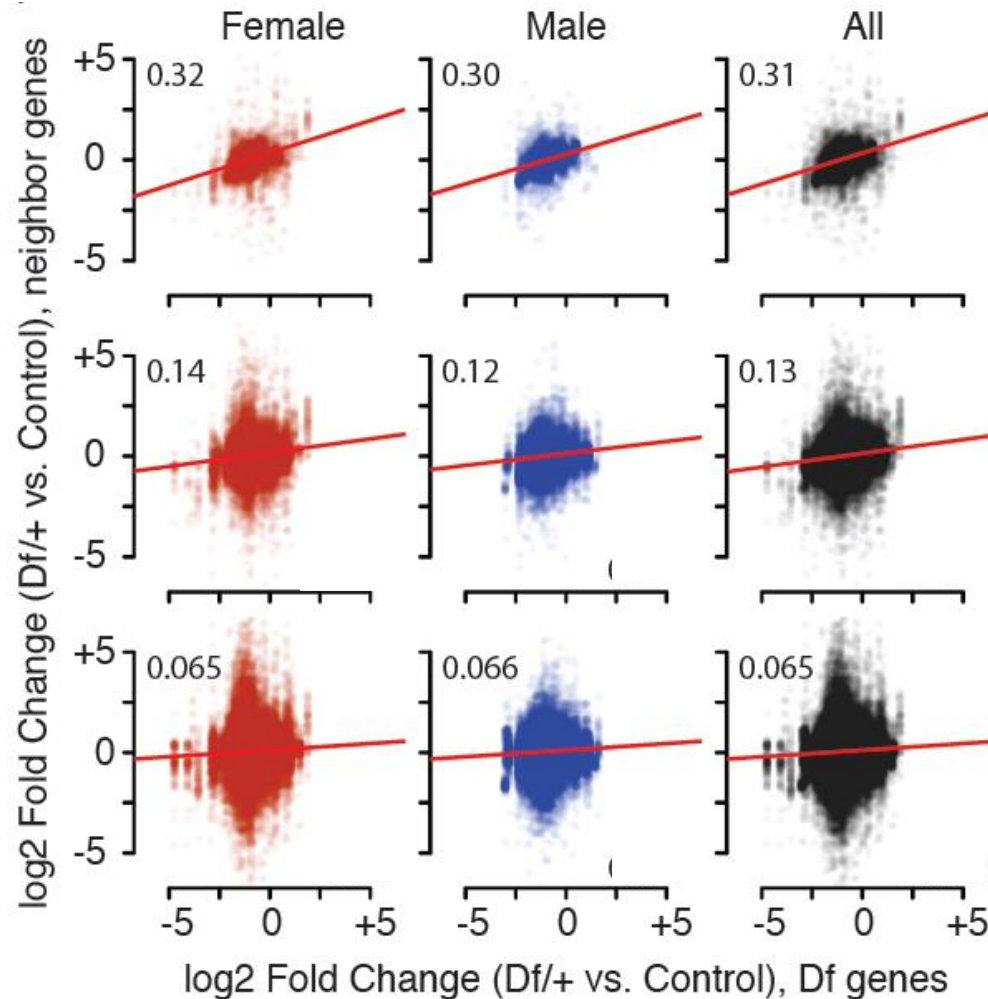
Expression of deficiency genes is reduced by nearly half  
(but not exactly and not always)

# What is the impact on network neighbors?





# Expression changes propagate through the network up to the second order neighbors

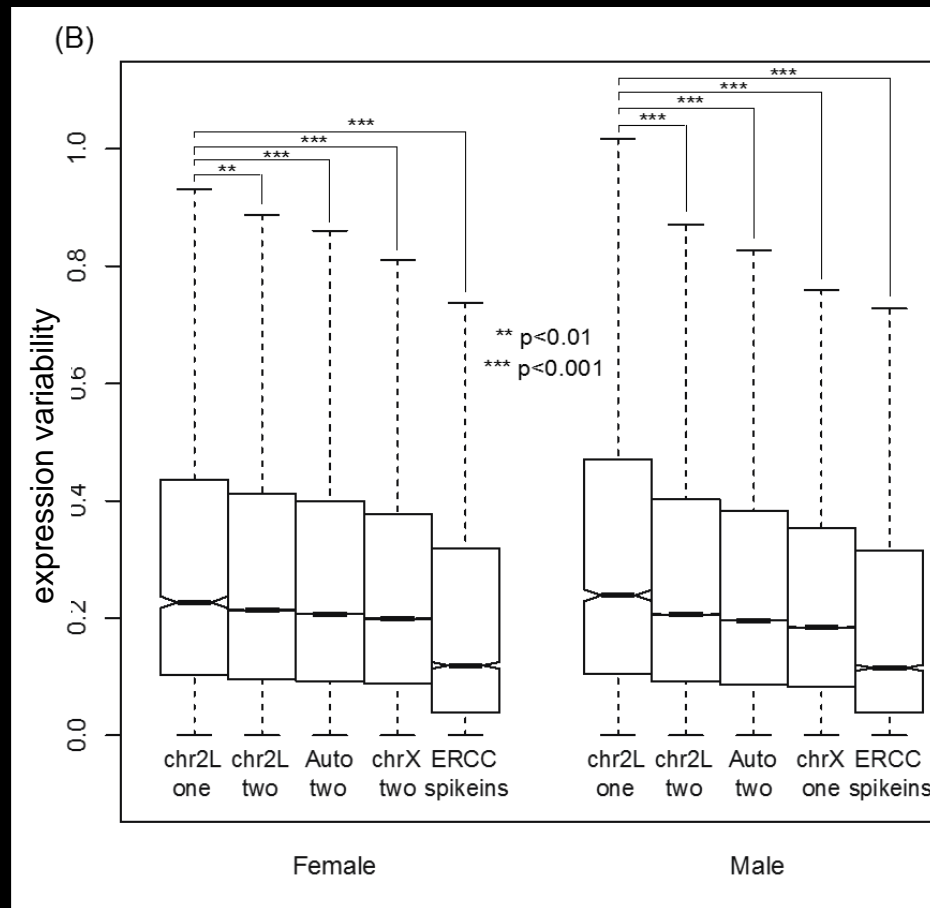


**First order neighbors**

**Second order neighbors**

**Third order neighbors**

# Expression variation of deficiency genes is higher than other genes

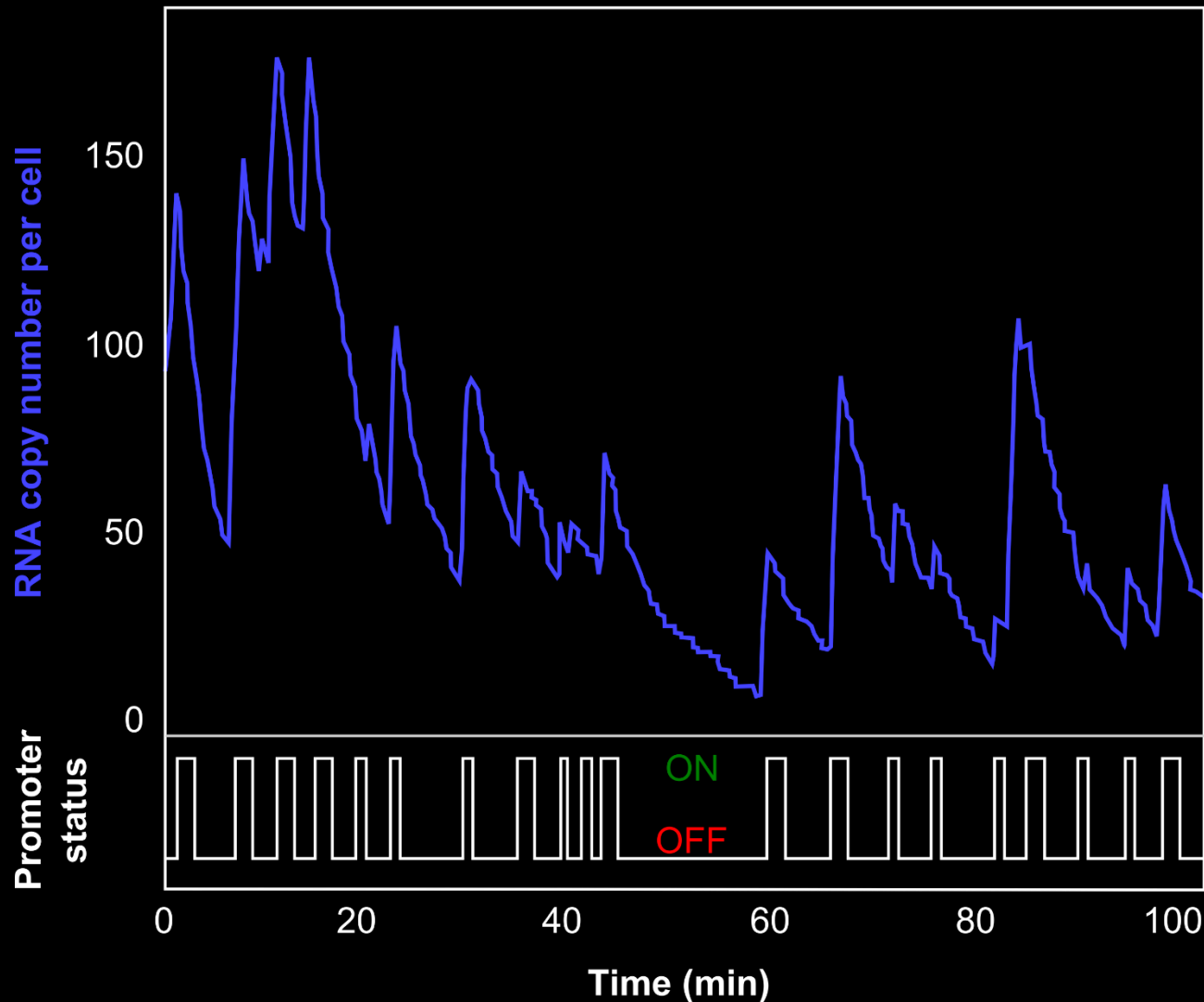


**Measurement of variability**

**$\Delta = \text{abs}(e1-e2) / \text{ave}(e1,e2)$**

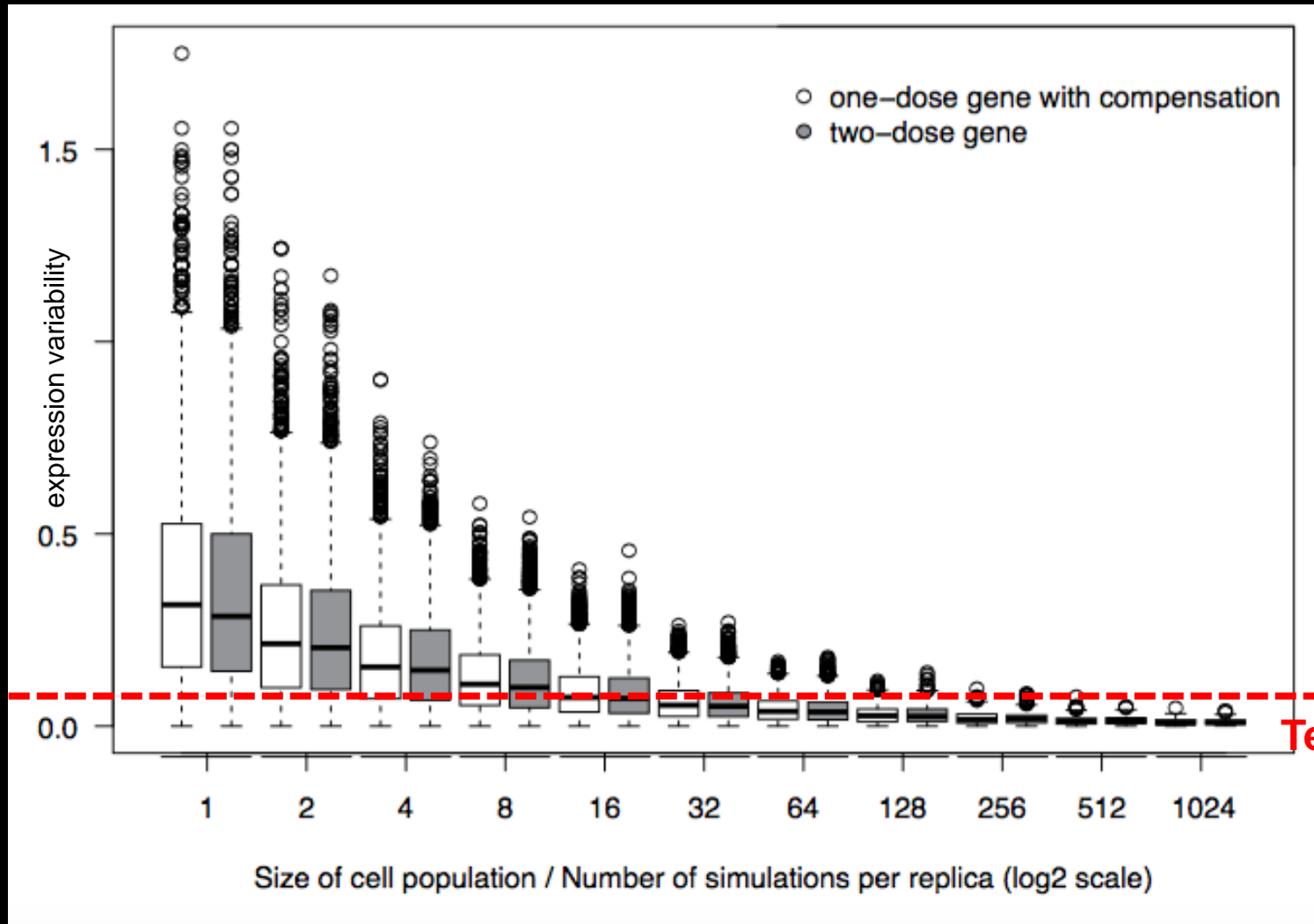
**$e1, e2$  – expression in experiment 1 and 2**

# Expression bursts as major contributor to single cell expression noise

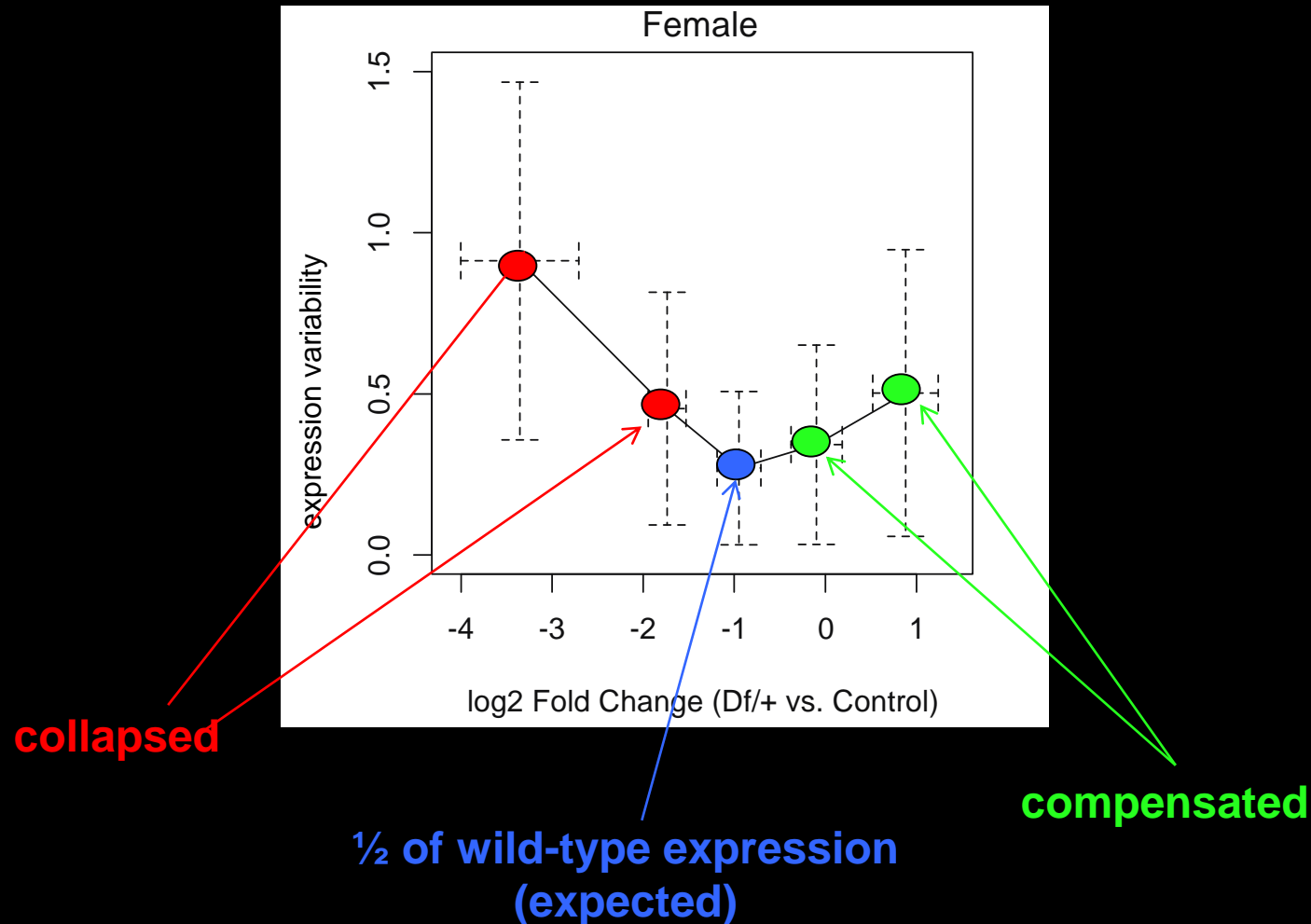


Prediction from the model – increased copy copy number reduces

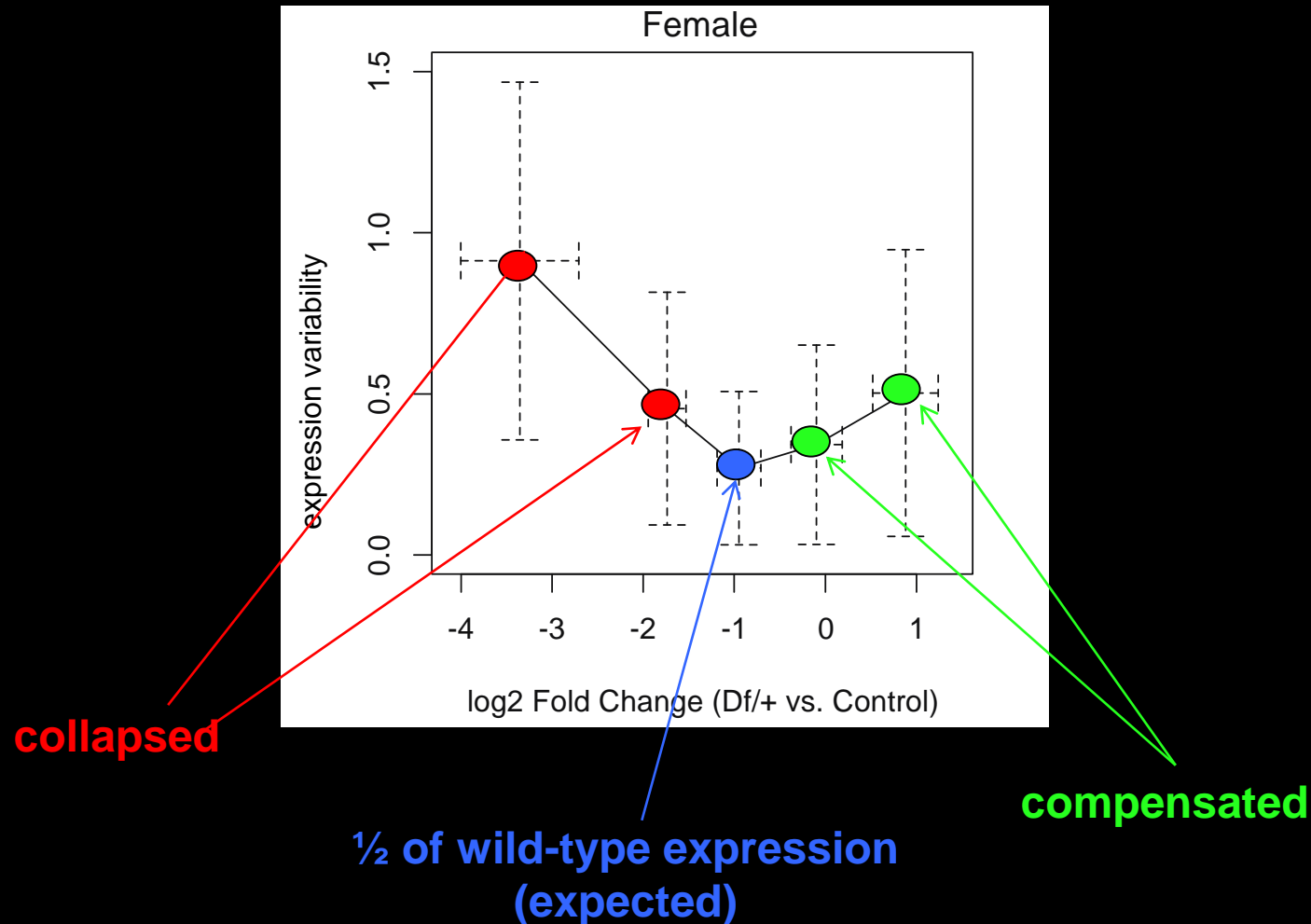
# Single cell model does not explain variability in expression of deficiency genes



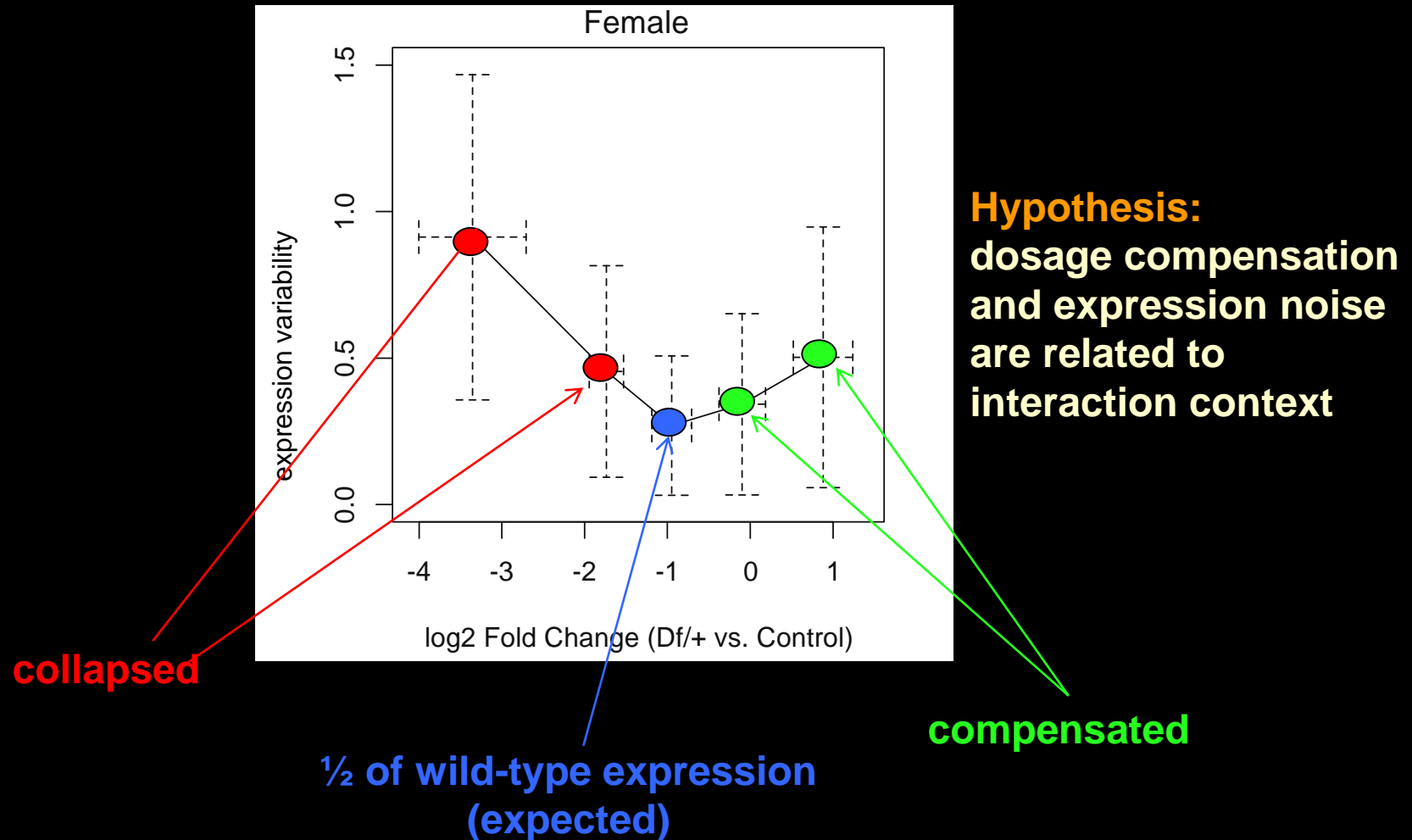
# Deficiency genes with compensated/collapsed expression are more noisy



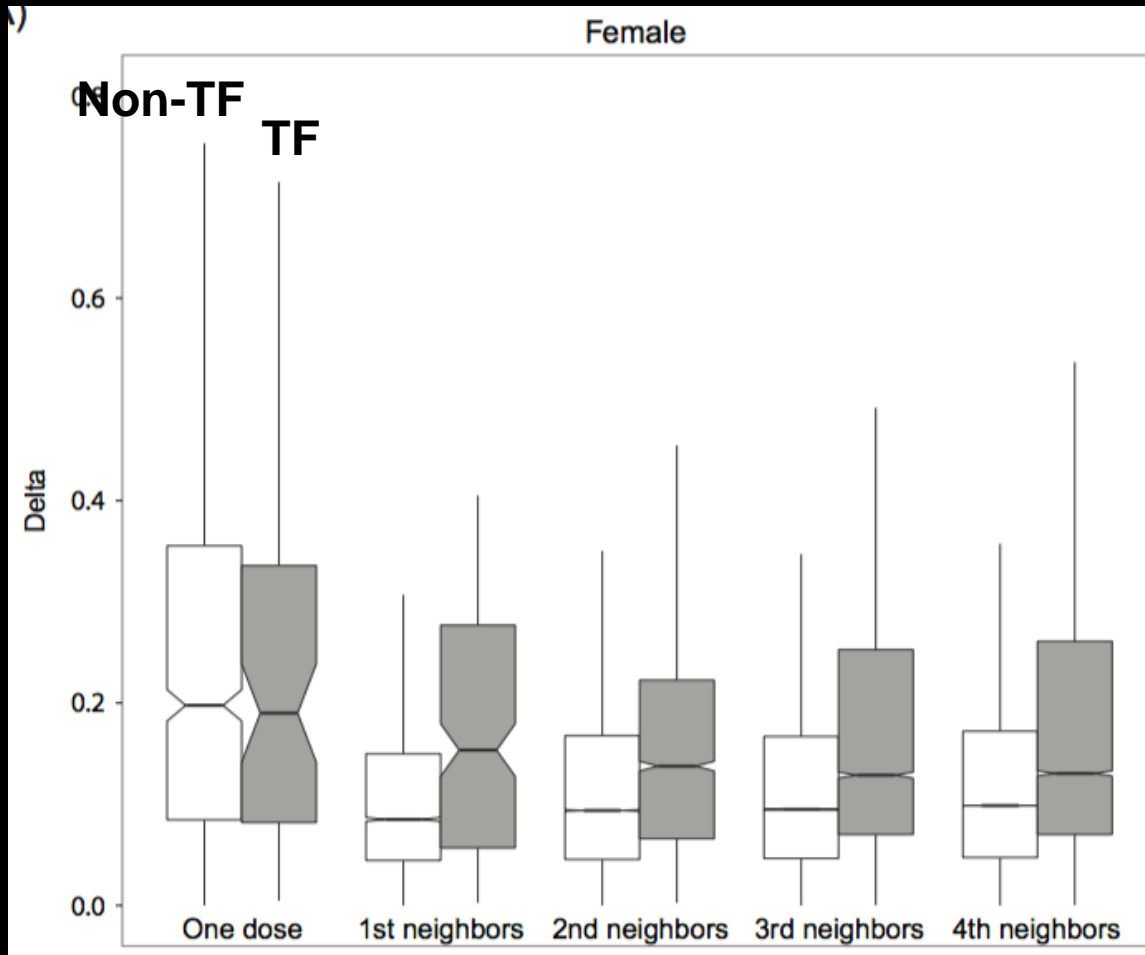
# Deficiency genes with compensated/collapsed expression are more noisy



# Deficiency genes with compensated/collapsed expression are more noisy



# Expression variations propagate across regulatory network





# Why this is relevant

## ARTICLE

### Identification of Small Exonic CNV from Whole-Exome Sequence Data and Application to Autism Spectrum Disorder

Christopher S. Poultney,<sup>1,2</sup> Arthur P. Goldberg,<sup>1,2,3</sup> Elodie Drapeau,<sup>1,2</sup> Yan Kou,<sup>1,4</sup>  
Hala Harony-Nicolas,<sup>1,2</sup> Yuji Kajiwara,<sup>1,2</sup> Silvia De Rubeis,<sup>1,2</sup> Simon Durand,<sup>1,2</sup> Christine Stevens,<sup>5</sup>  
Karola Rehnström,<sup>6,7</sup> Aarno Palotie,<sup>5,6</sup> Mark J. Daly,<sup>5,8</sup> Avi Ma'ayan,<sup>4</sup> Menachem Fromer,<sup>2,9</sup>  
and Joseph D. Buxbaum<sup>1,2,3,9,10,\*</sup>



#### ANNUAL REVIEWS **Further**

Click [here](#) for quick links to  
Annual Reviews content online,  
including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

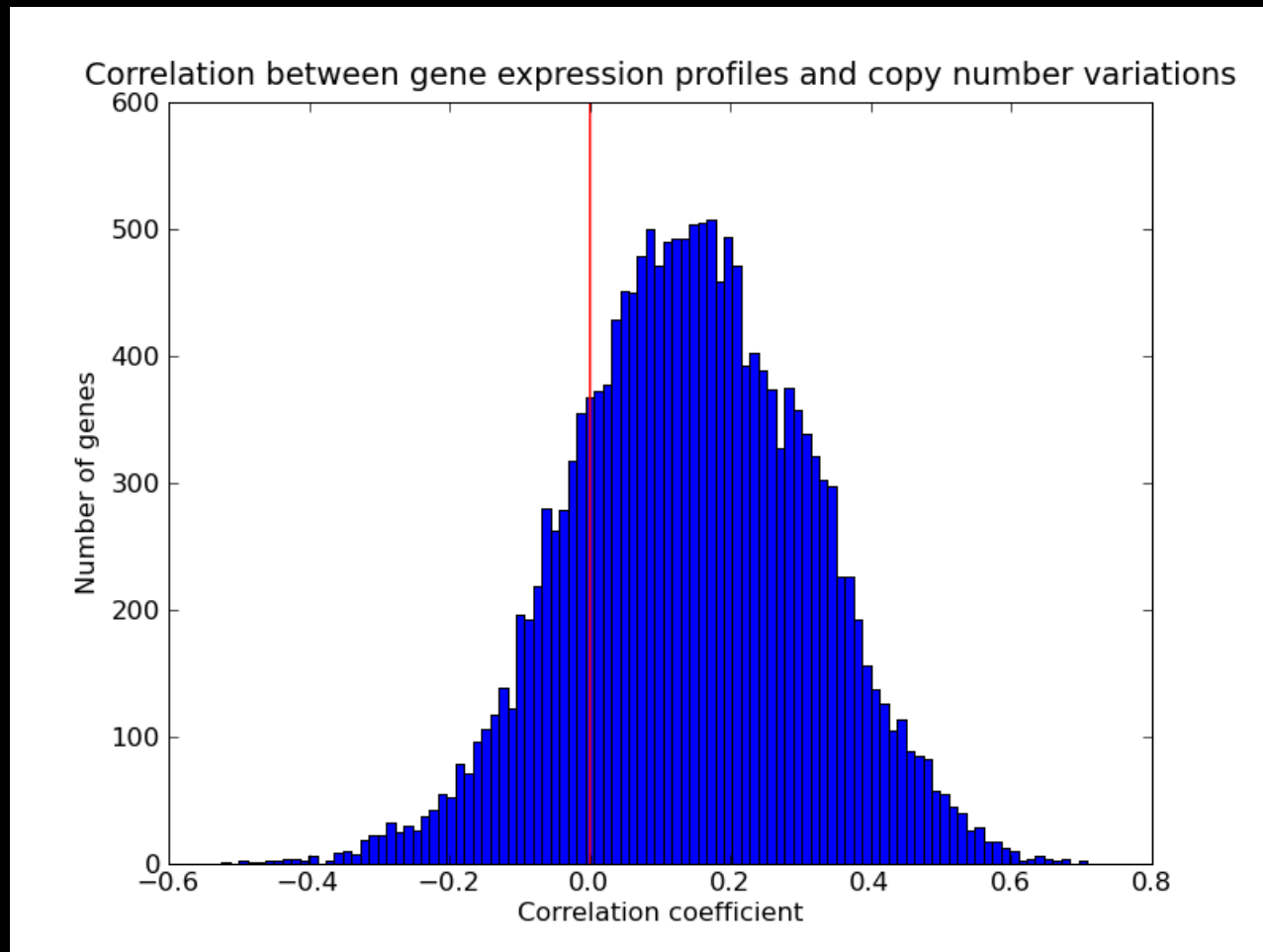
### Human Copy Number Variation and Complex Genetic Disease

Santhosh Girirajan, Catarina D. Campbell,  
and Evan E. Eichler

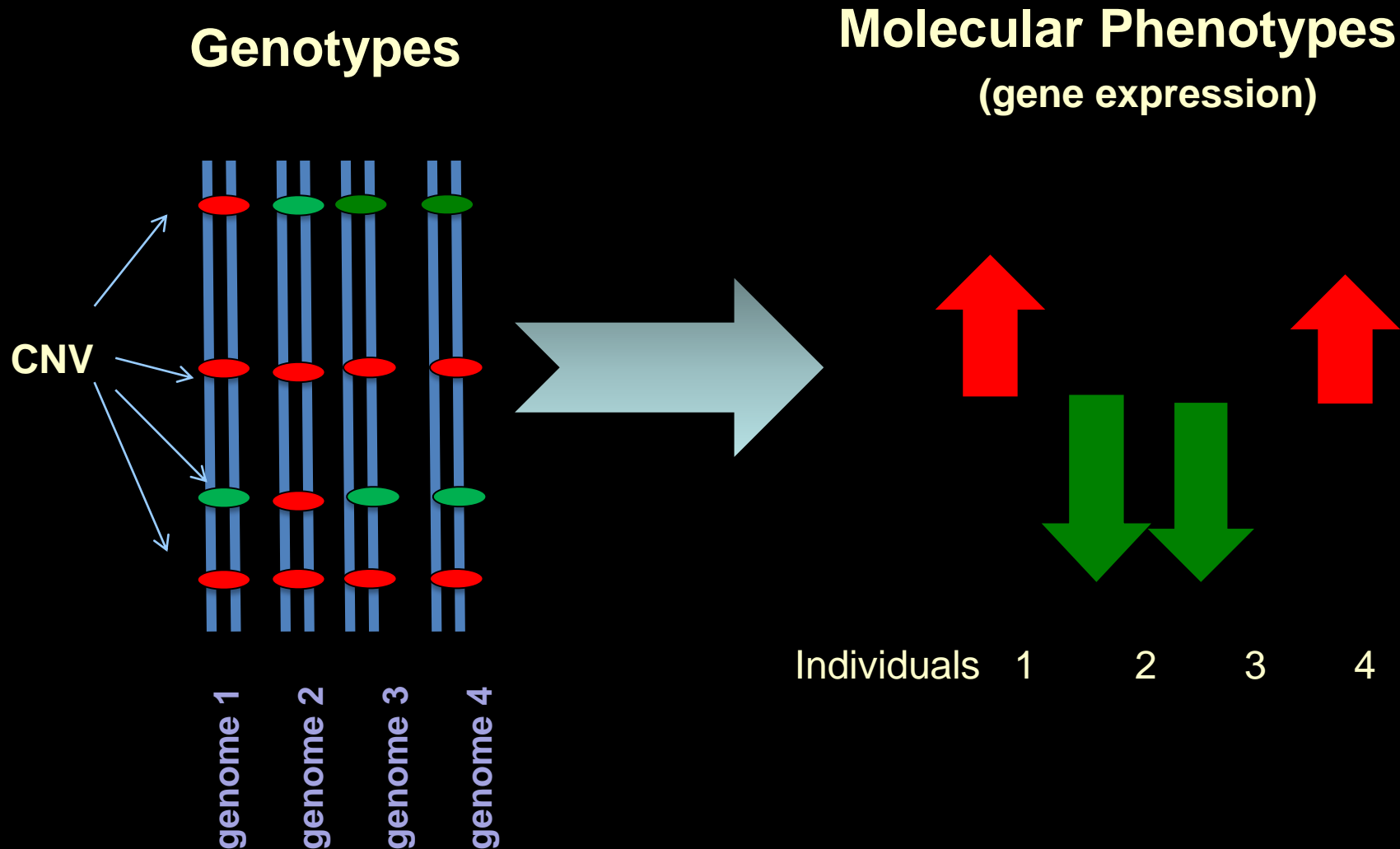
Department of Genome Sciences and Howard Hughes Medical Institute, University of  
Washington, Seattle, Washington 98195; email: [sangi@u.washington.edu](mailto:sangi@u.washington.edu),  
[cdcamp@u.washington.edu](mailto:cdcamp@u.washington.edu), [eee@gs.washington.edu](mailto:eee@gs.washington.edu)

# **Copy number variations and propagation of expression changes in cancer**

# Positive correlation between gene copy number and gene expression in cancer (GBM)



# Using propagation of expression perturbation to study genotype-phenotype relation

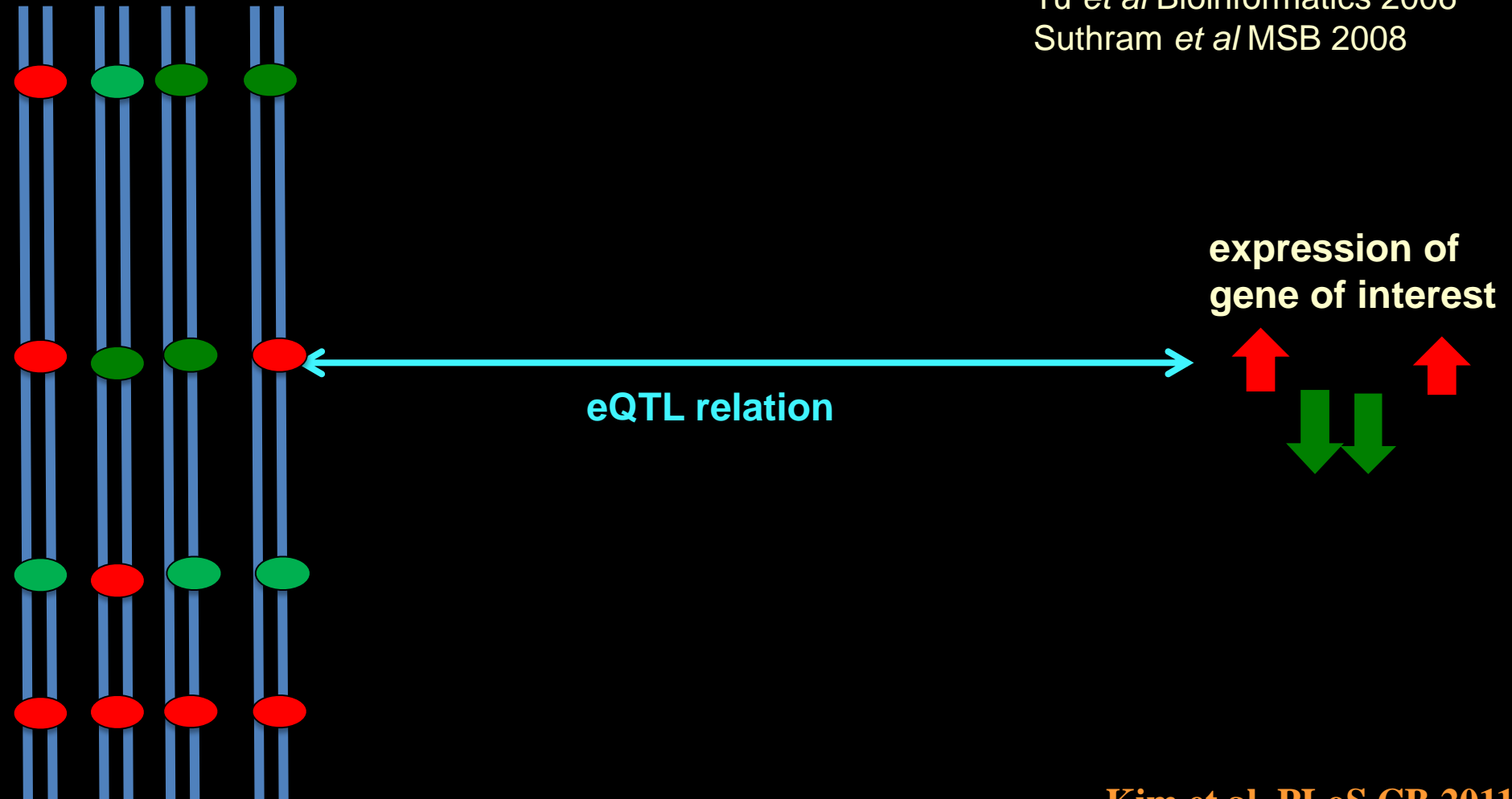


# The eQTL-net method

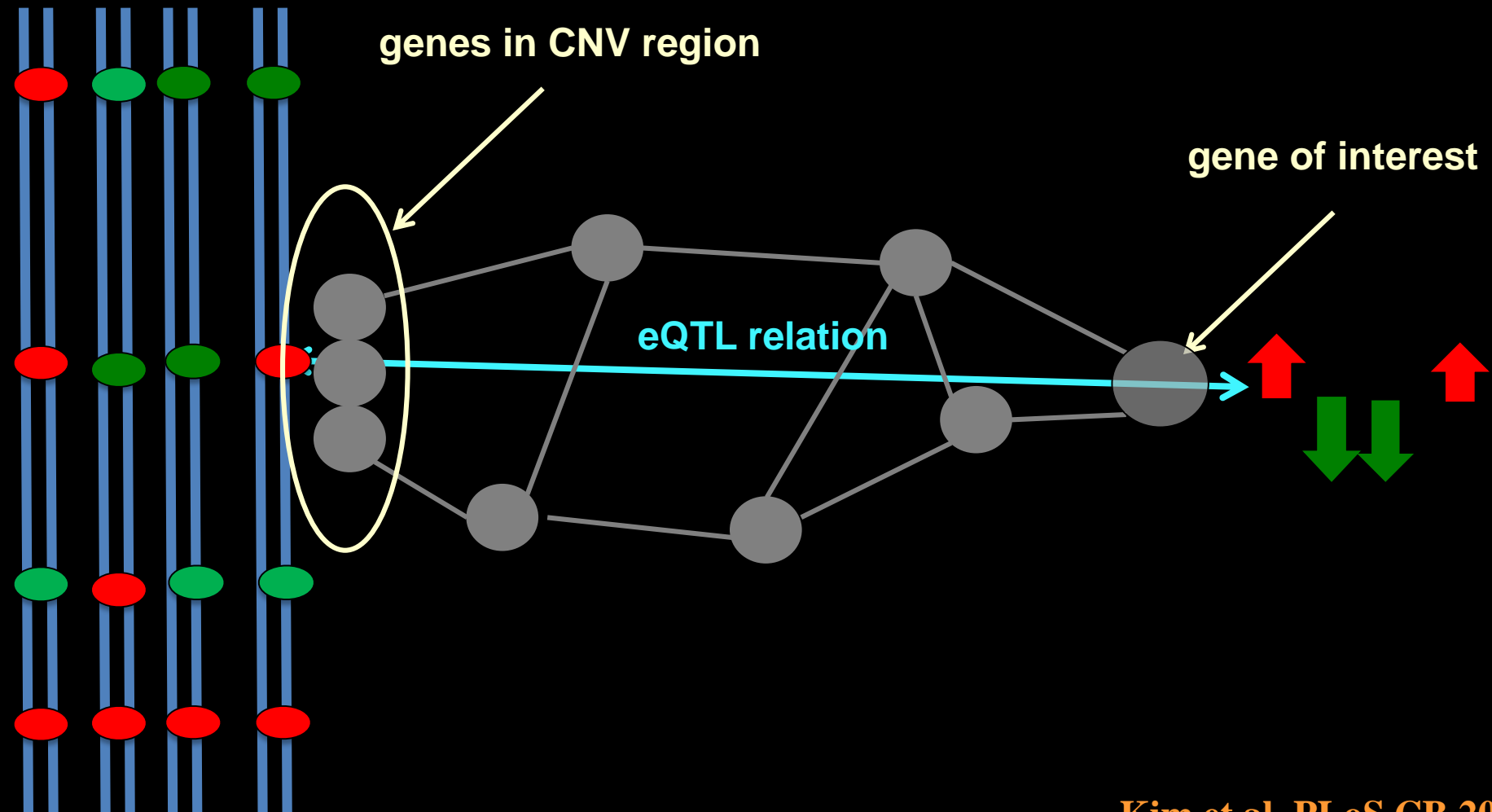
Motivated in part by:

Tu *et al* Bioinformatics 2006

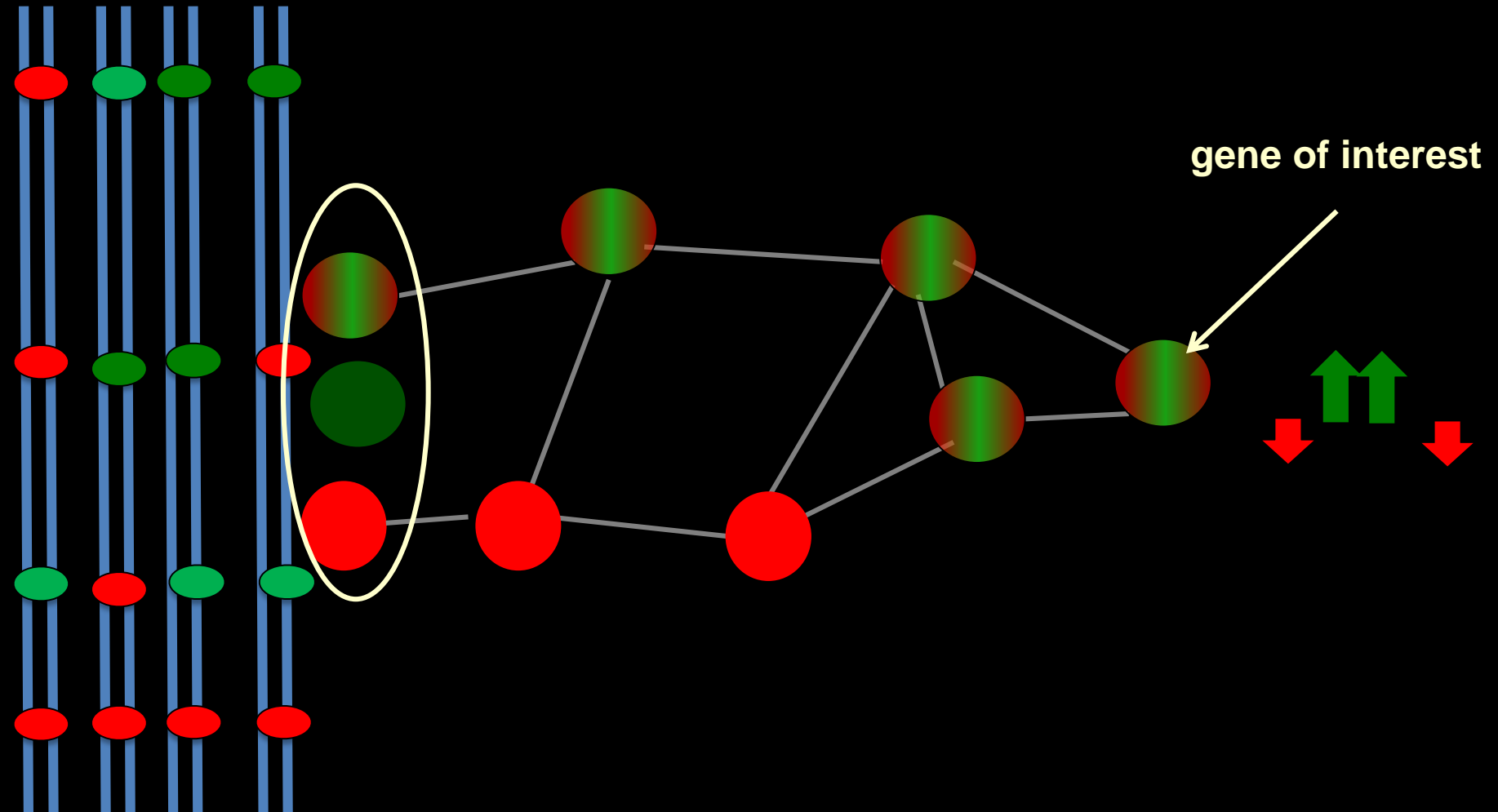
Suthram *et al* MSB 2008



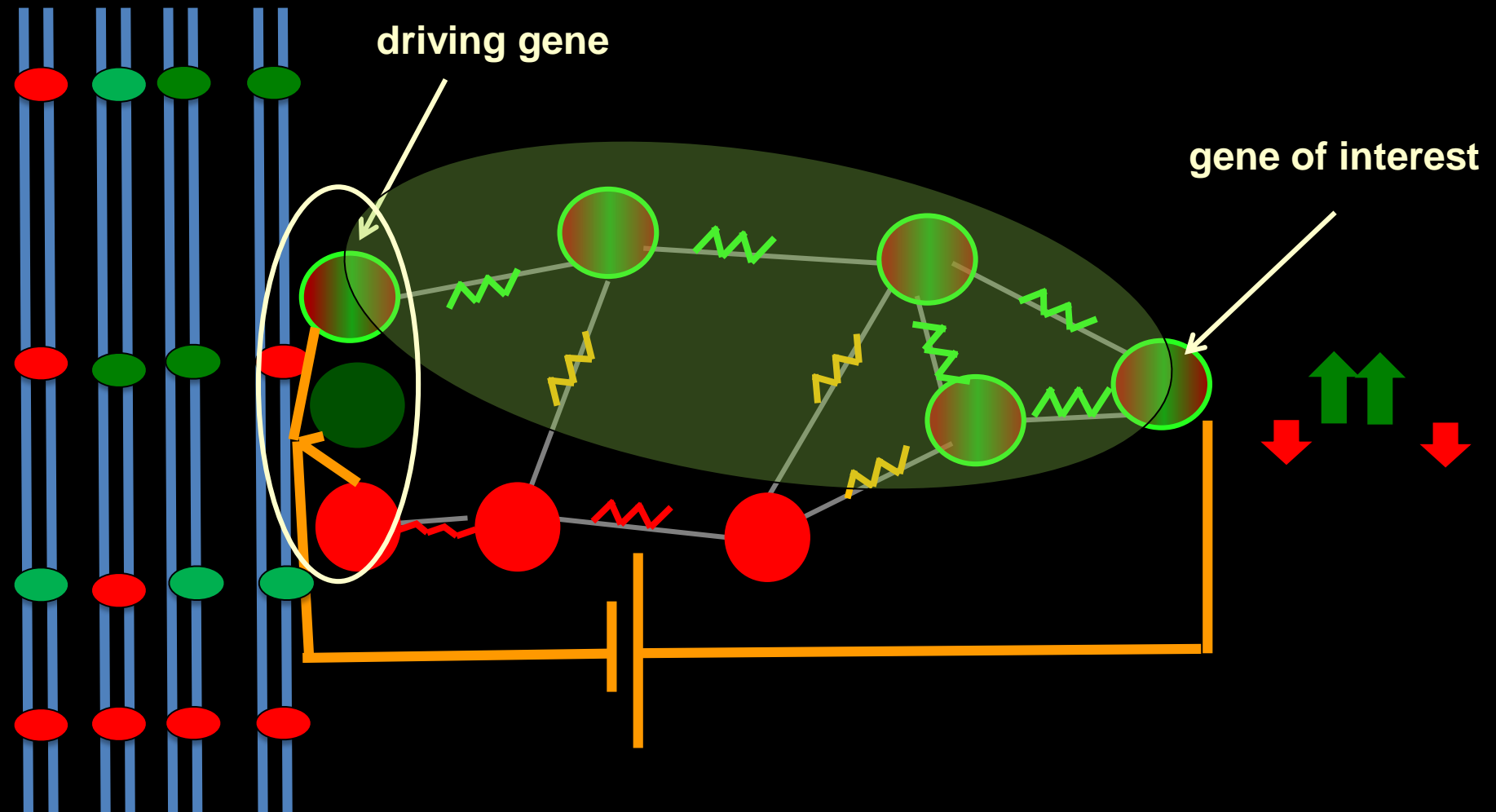
# Explaining eQTL associations by propagation of expression changes



# Explaining eQTL associations by propagation of expression changes

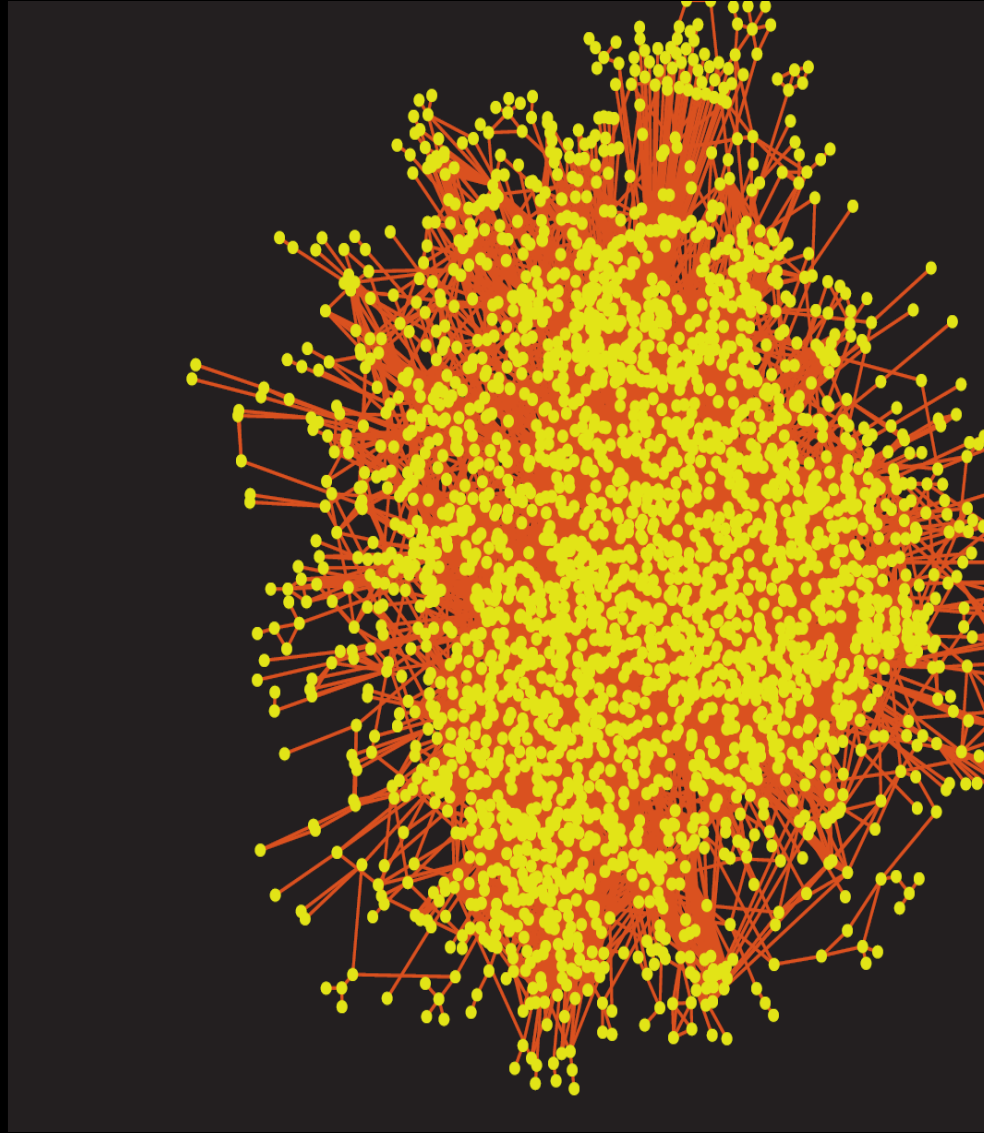


## Identification of information propagation pathways and driver genes

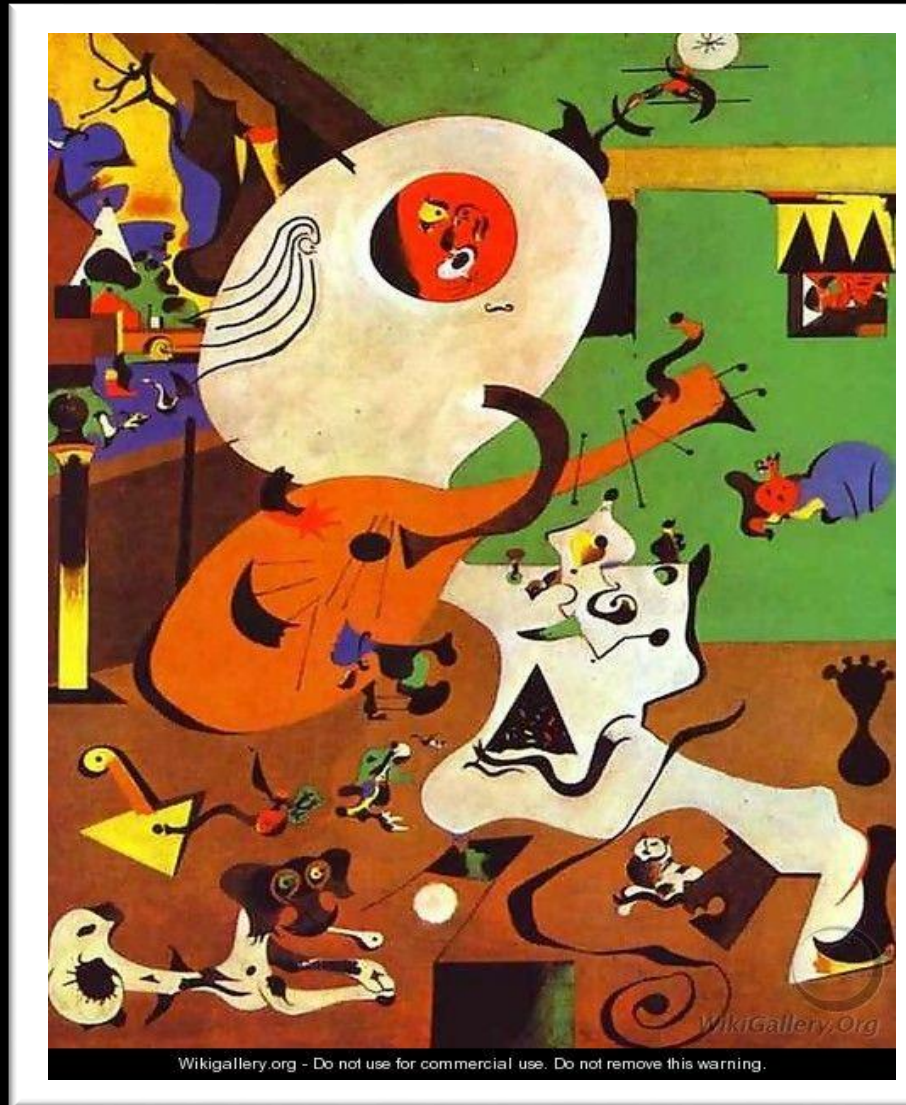




# How to interpret these pathway in the context of throughput interaction networks?



# How to interpret this painting?

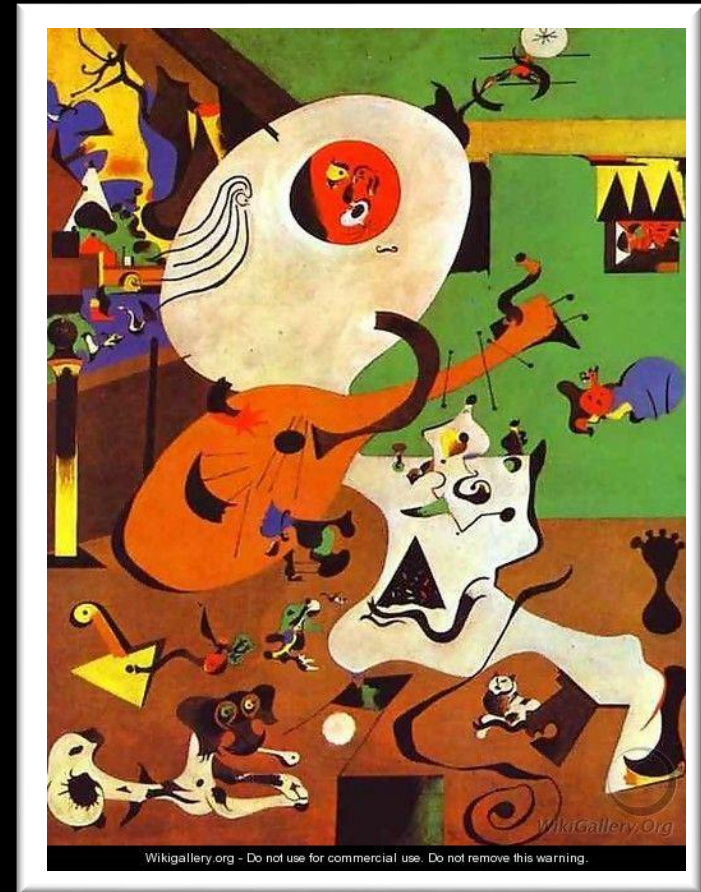


Dutch Interior 1, Joan Miro' (1893–1983) Museum of Modern Art, New York  
© 2012 Successió Miró / Artists Rights Society (ARS), New York / ADAGP, Paris used with ARS permission).

# Details are perturbed but relationships remain



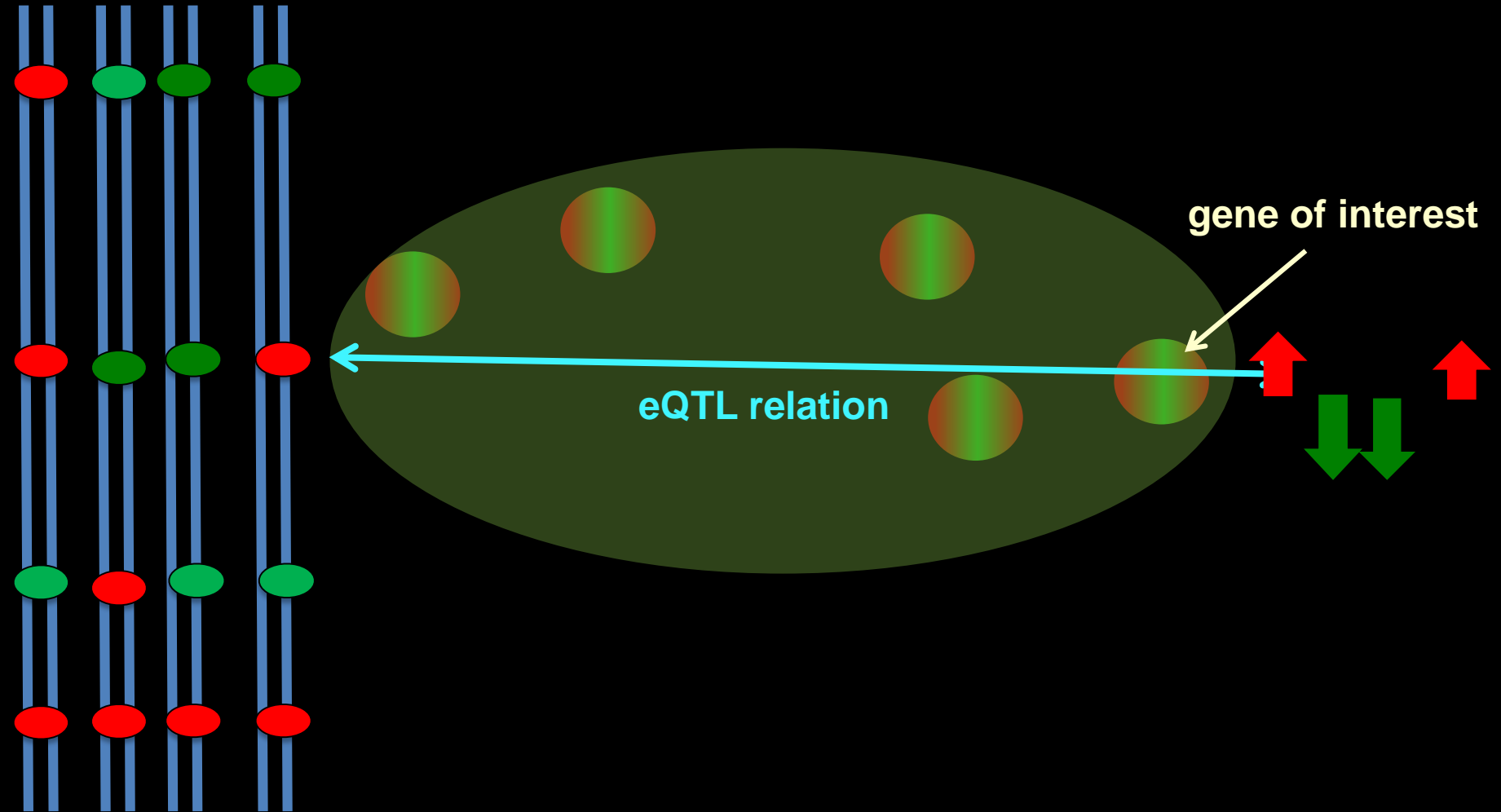
**The Lute Player, Hendrick Maertensz Sorgh (1610-1670),  
Rijksmuseum, Amsterdam  
(public domain)**



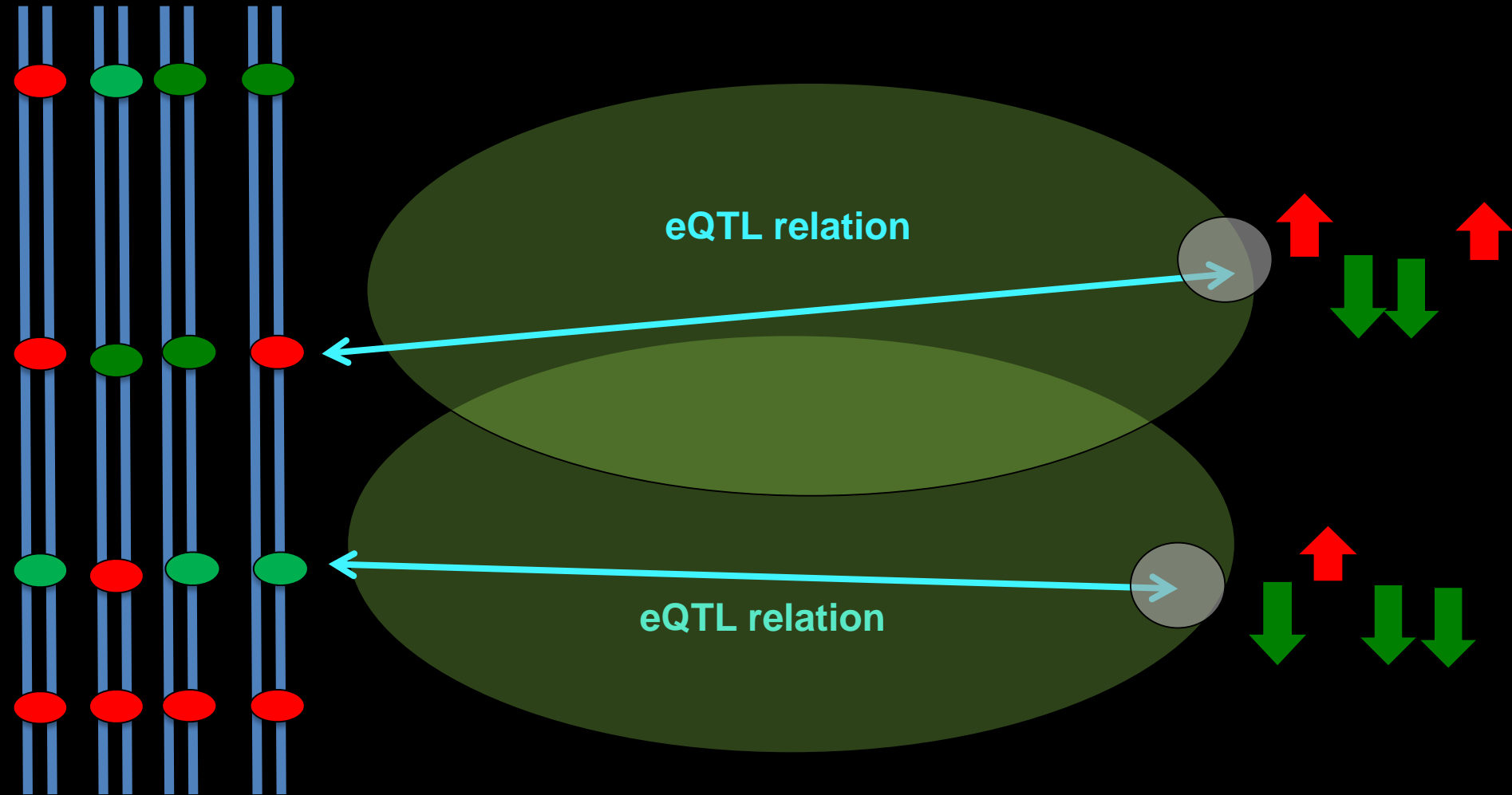
**Dutch Interior 1, Joan Miro' (1893-1983)  
Museum of Modern Art, New York**

© 2012 Successió Miró / Artists Rights Society (ARS), New York / ADAGP, Paris  
(used with ARS permission).

# Considering information propagation pathway as a bag of genes



Typically we use many genes as a molecular phenotype- repeat the process for all of them





# Results

## Driving Copy Number alterations

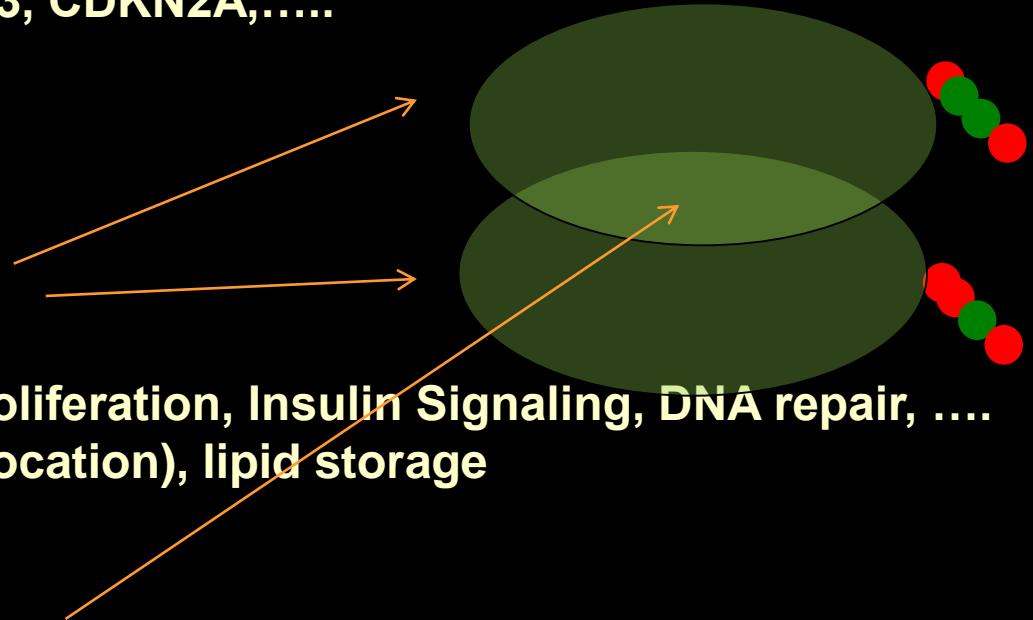
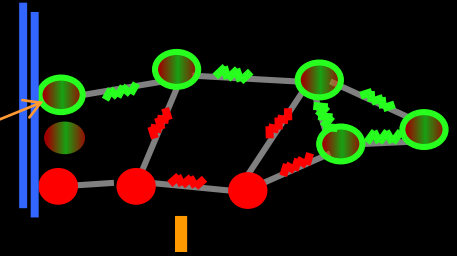
EGFR, PTEN, RB1, GBAS, TP53, CDKN2A,.....

## Reoccurring pathways

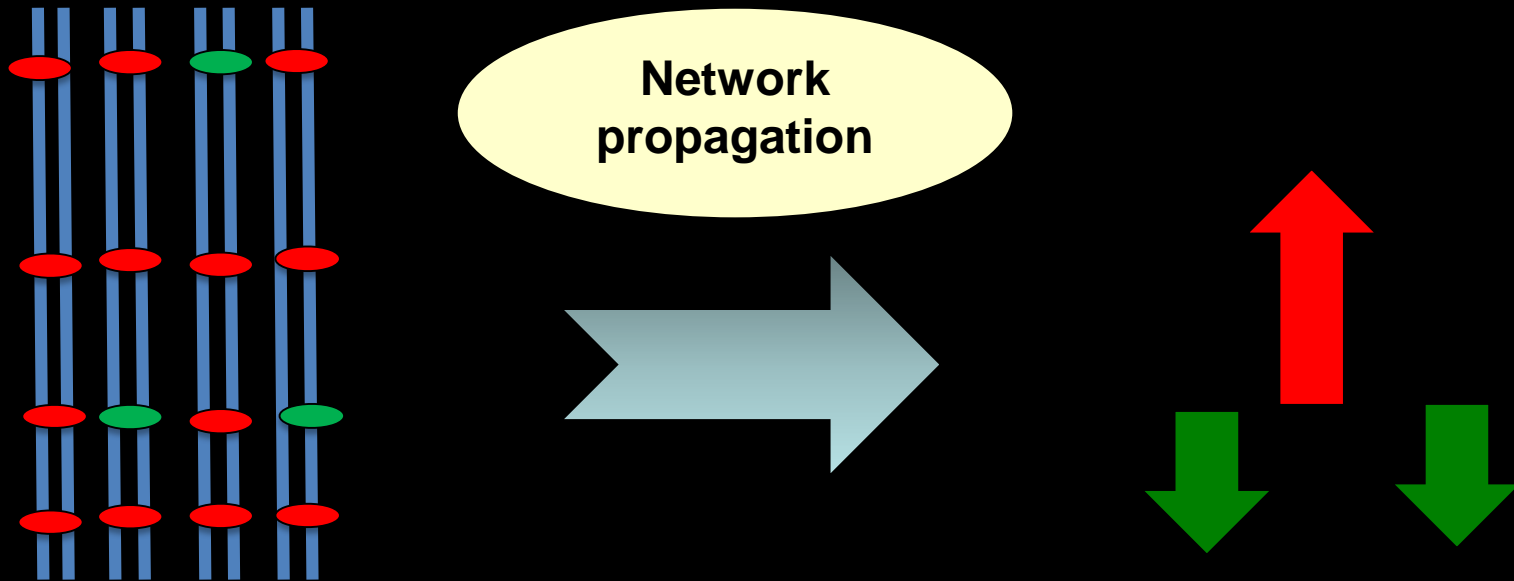
Cell cycle, EGFR, RAS, Cell proliferation, Insulin Signaling, DNA repair, ....  
Splicing, SMAD (nuclear translocation), lipid storage

## Hubs

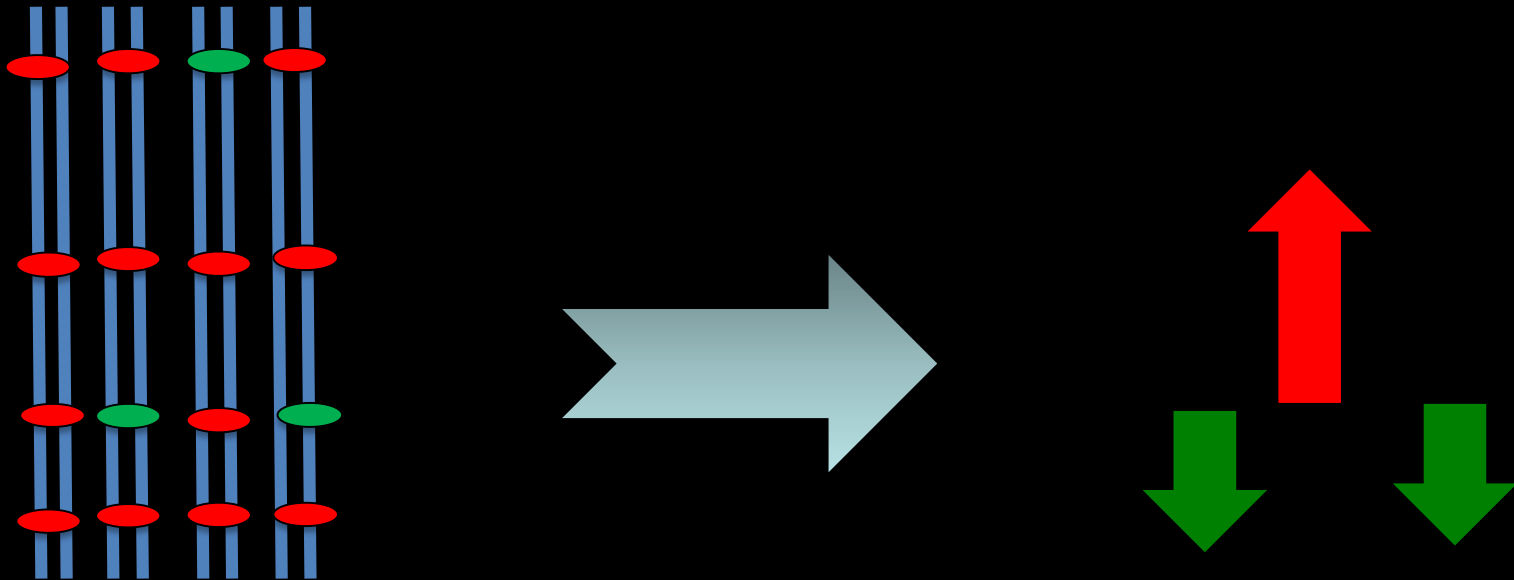
Myc, E2F1, CREBBP, SP1, Jun,...



# Utilizing Networks for Understanding Genotype-Phenotype relations

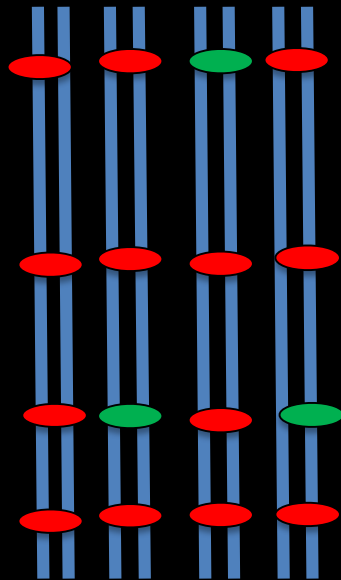


**What can we do about perturbations that do not necessarily cause expression change in network neighbors?**





# Finding mutated/dysregulated subnetworks



**Main principle:**

If different perturbations have similar effects the perturbations should be related – belong to the same subnetwork

**Challenge:**

Cancer heterogeneity

mutated /  
dysregulated  
subnetworks

# Module Cover Approach

Motivated in part by: Ulitsky et al 2008

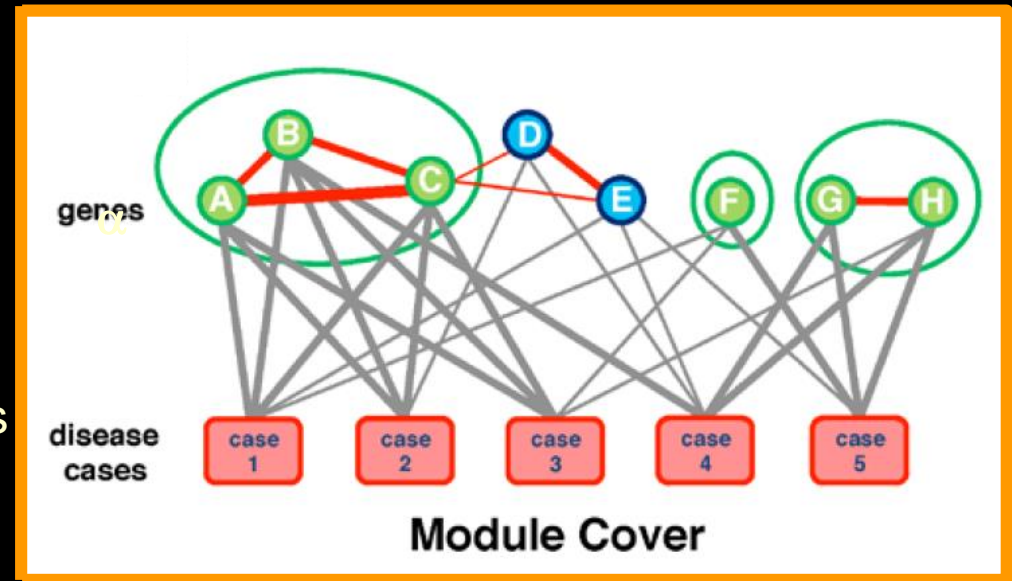
## Optimization problem:

Find minimal cost set of modules so that each disease case is covered at least  $k$  times

**Cost** is determined by:

↓ A similarity of genes within modules (application dependent)

↑ number of modules



**Advantage:** different patients can be covered by different subnetworks

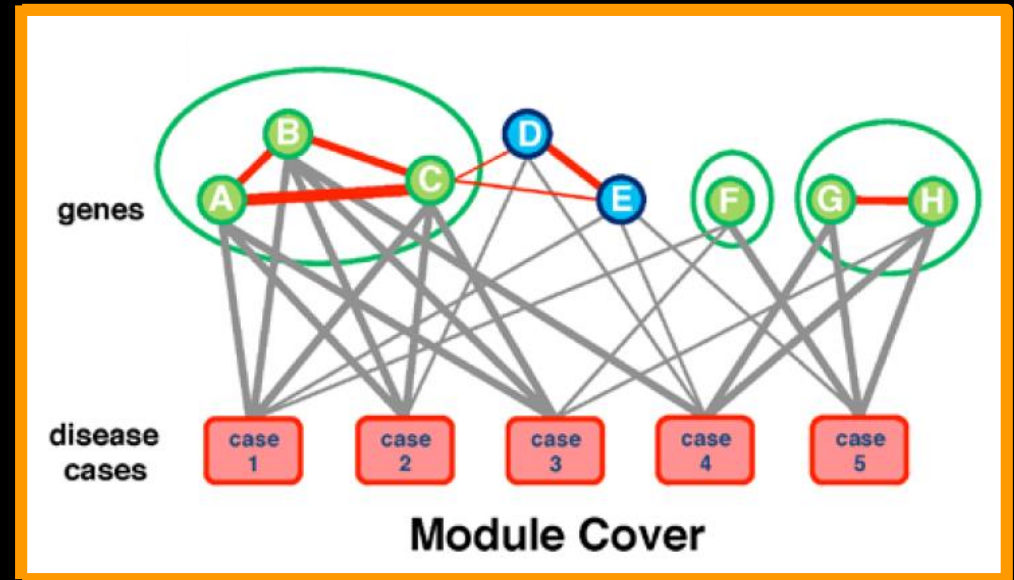
Kim et al. 2013

# Application 1: Glioblastoma Analysis

**Cost** is a function of:

- ↓ Similarity within modules
- Distance in network
  - Expression similarity (similarity of eQTL profiles)

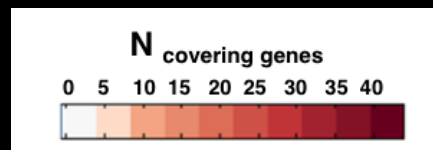
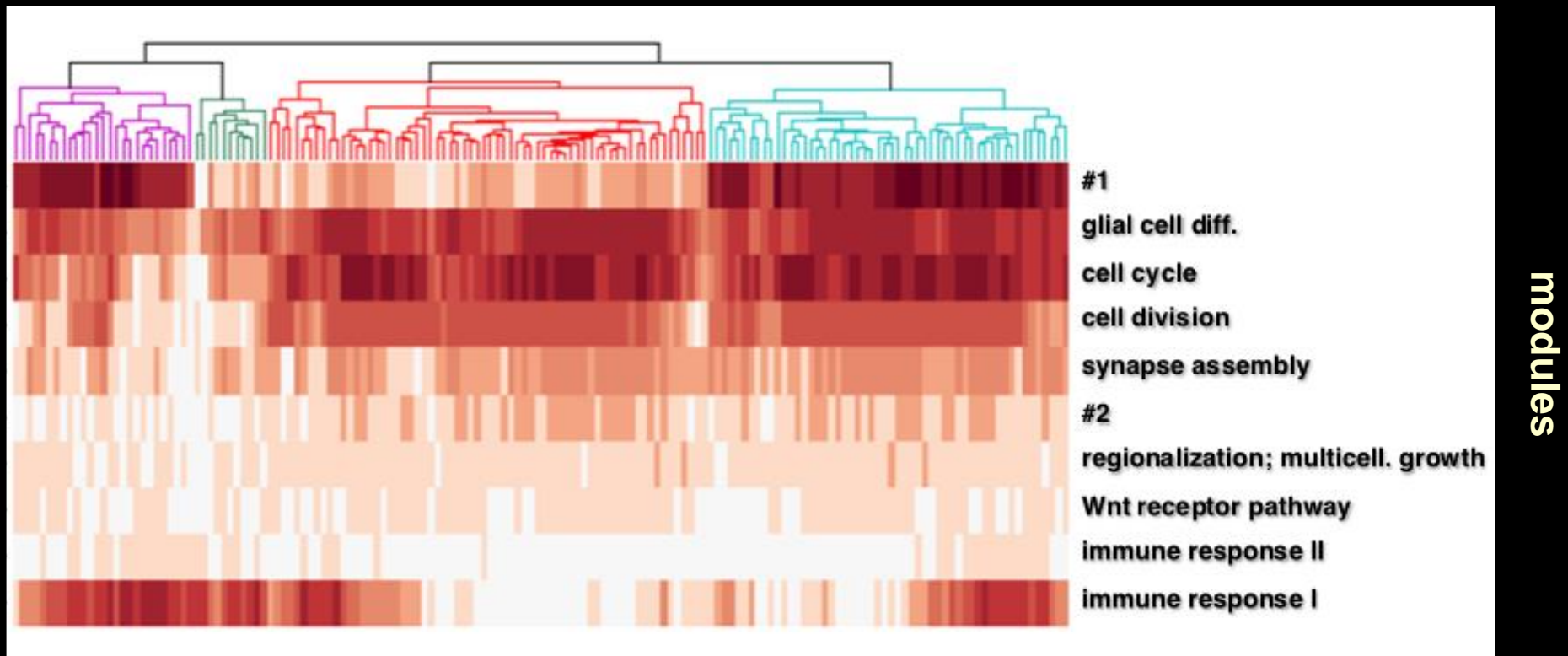
↑ number of modules  
(parameterized penalty)



**Advantage:** different patients can be covered by different subnetworks

# Modules capture patients heterogeneity/subtypes

cases



## Application 2: Extension to mutual exclusivity

# Mutual exclusivity of cancer drivers

*Thomas et al 2007*

*Ciriello, et al., 2012;*

*Vandin, et al., 2012;*

*Szczurek et.al , 2014, 2015*

*Leiserson, et al., Vandin et al. 2013,2014,2015;*

*Kim et al. 2015*

*Constantinescu et al. 2015*

patients



mutations in gene 1

Mutations in gene 2

## Proposed explanations

- any of the two drivers alone gives sufficient growth advantage
- negative genetic interactions between drivers

## Interesting property

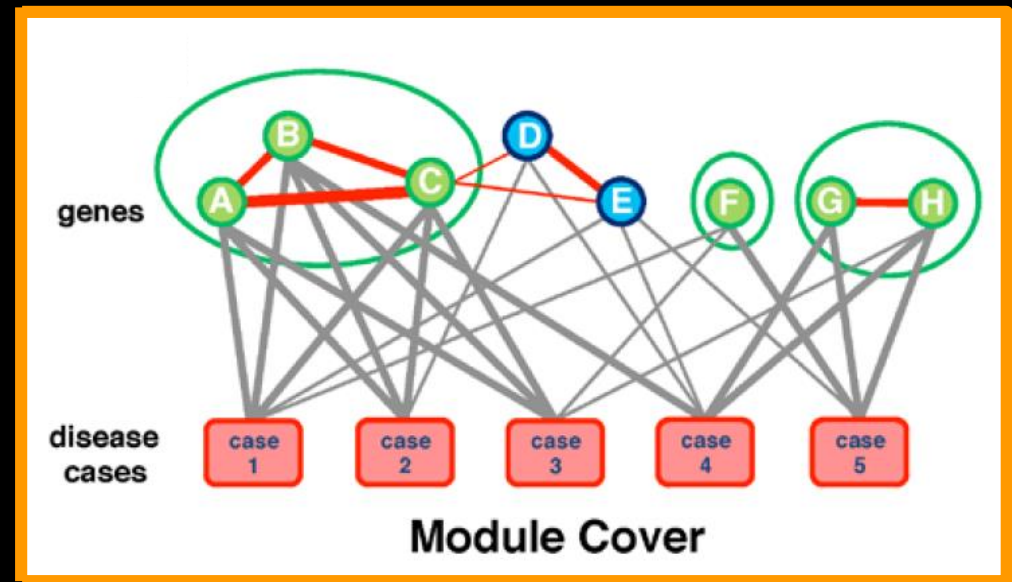
Mutual exclusive pairs are often in the same pathway

# Application 2: Subnetworks dysregulated across many cancer types

Cost is a function of:

- ↓ Similarity measure
- Distance in network
  - Mutual Exclusivity Score

↑ number of (parameterized penalty)



# Defining Mutual exclusivity in PanCancer setting

patients



patients

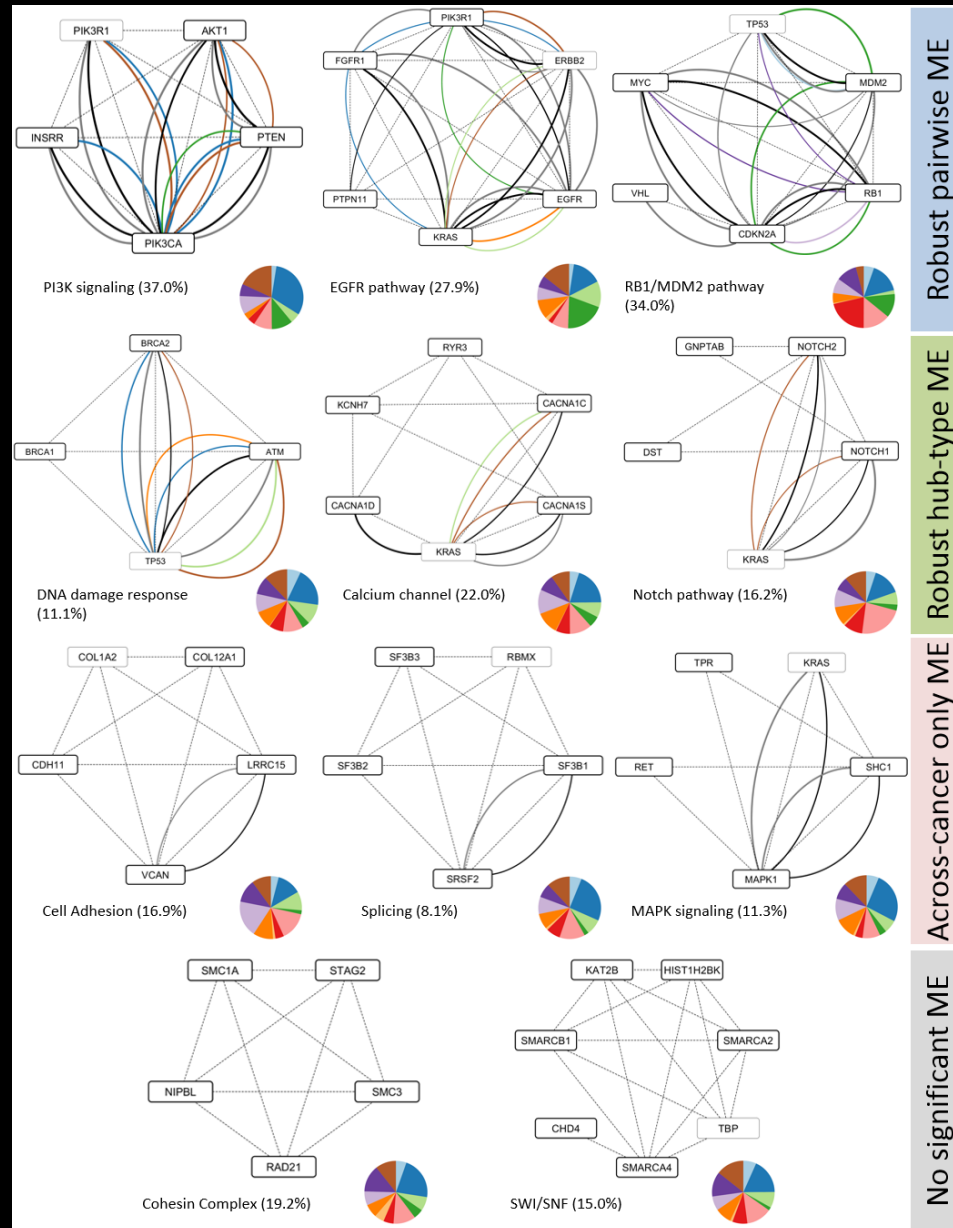


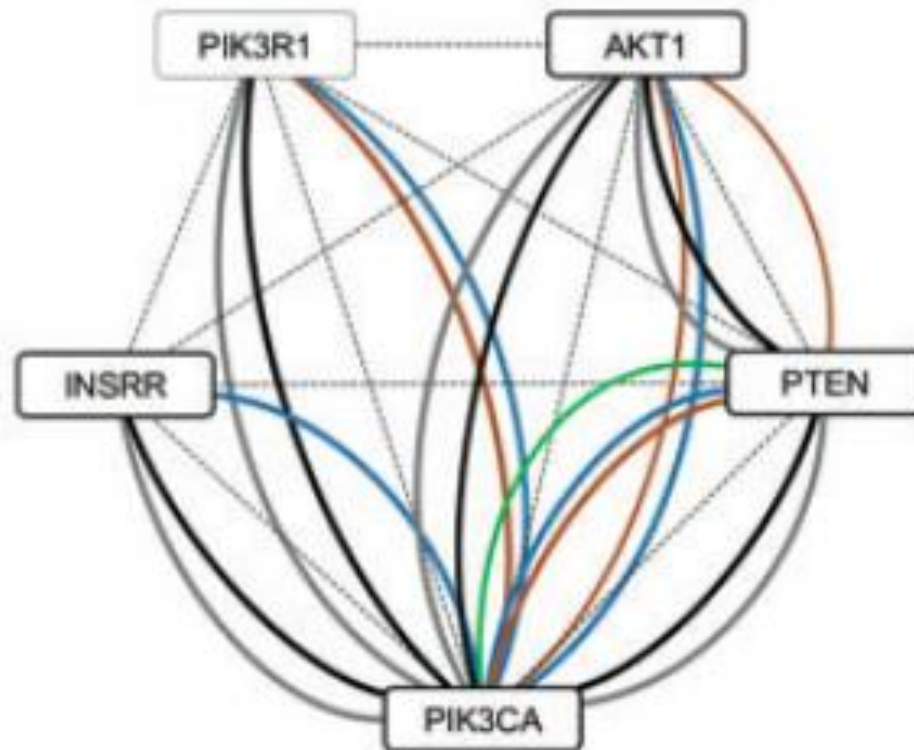
Kim et al. IMSB 2015 – classification of ME types in context of PanCancer and properties of different ME types



# MEMCover (Mutual Exclusivity Module Cover)

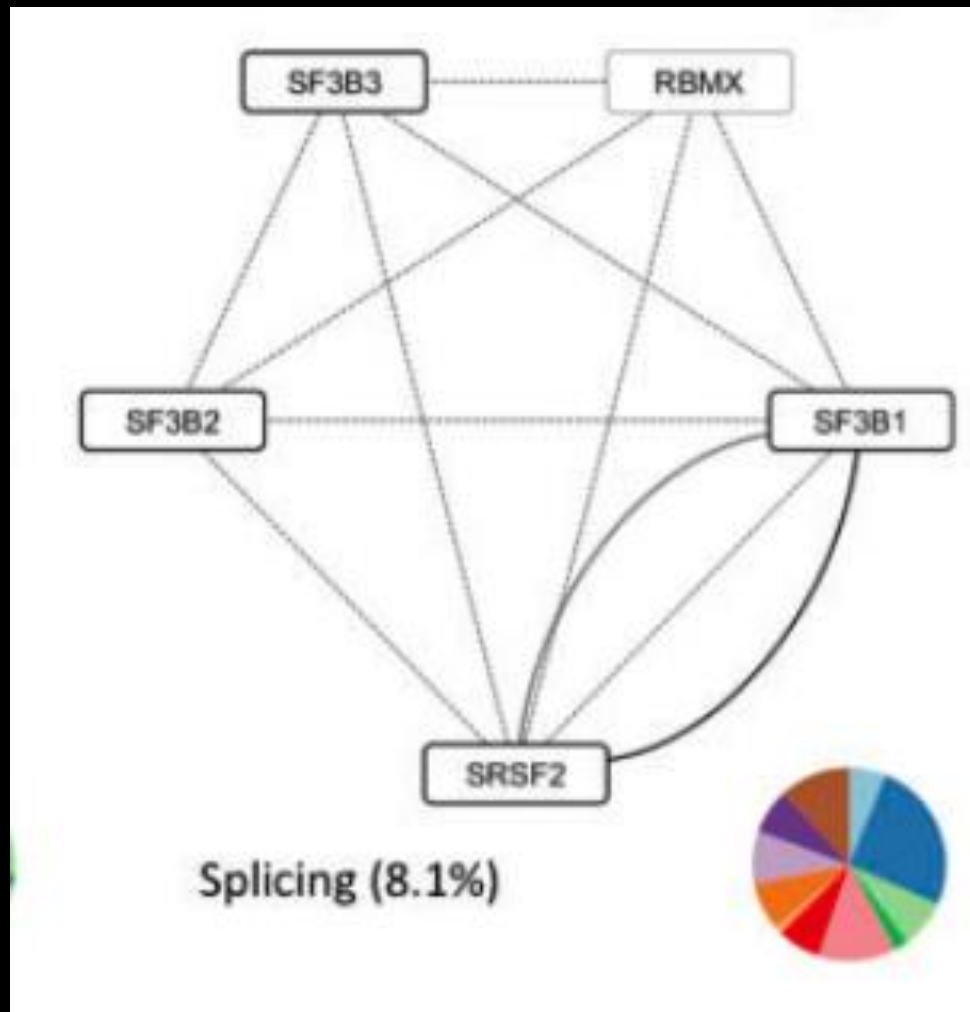
## Finds subnetworks dysregulated across cancer types





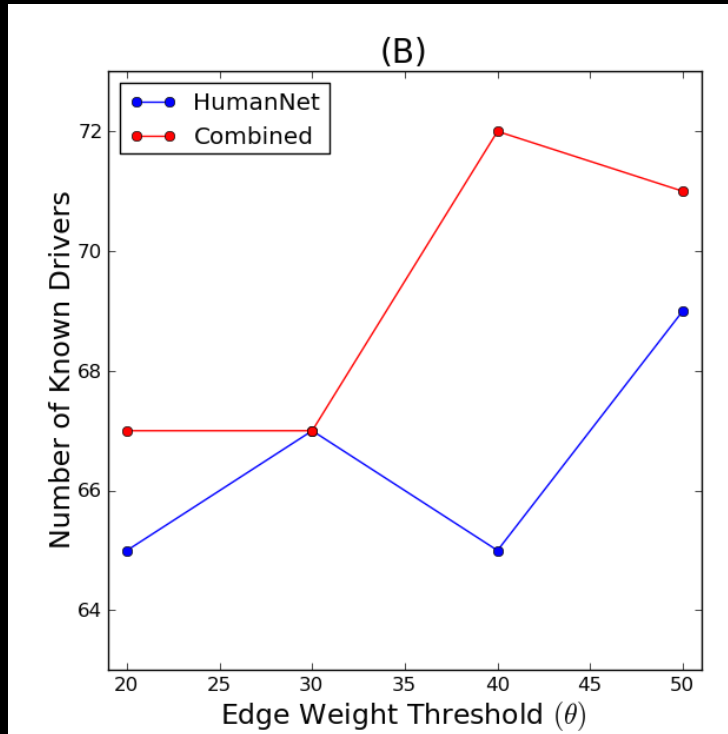
PI3K signaling (37.0%)



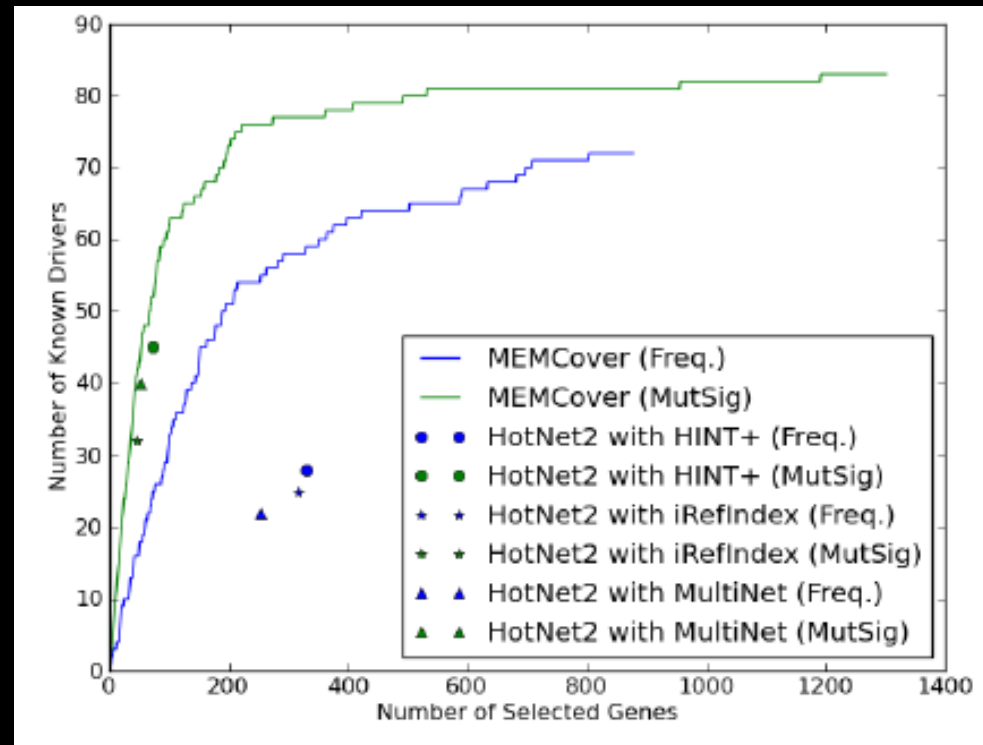


# MEMCover modules are enriched in cancer drivers

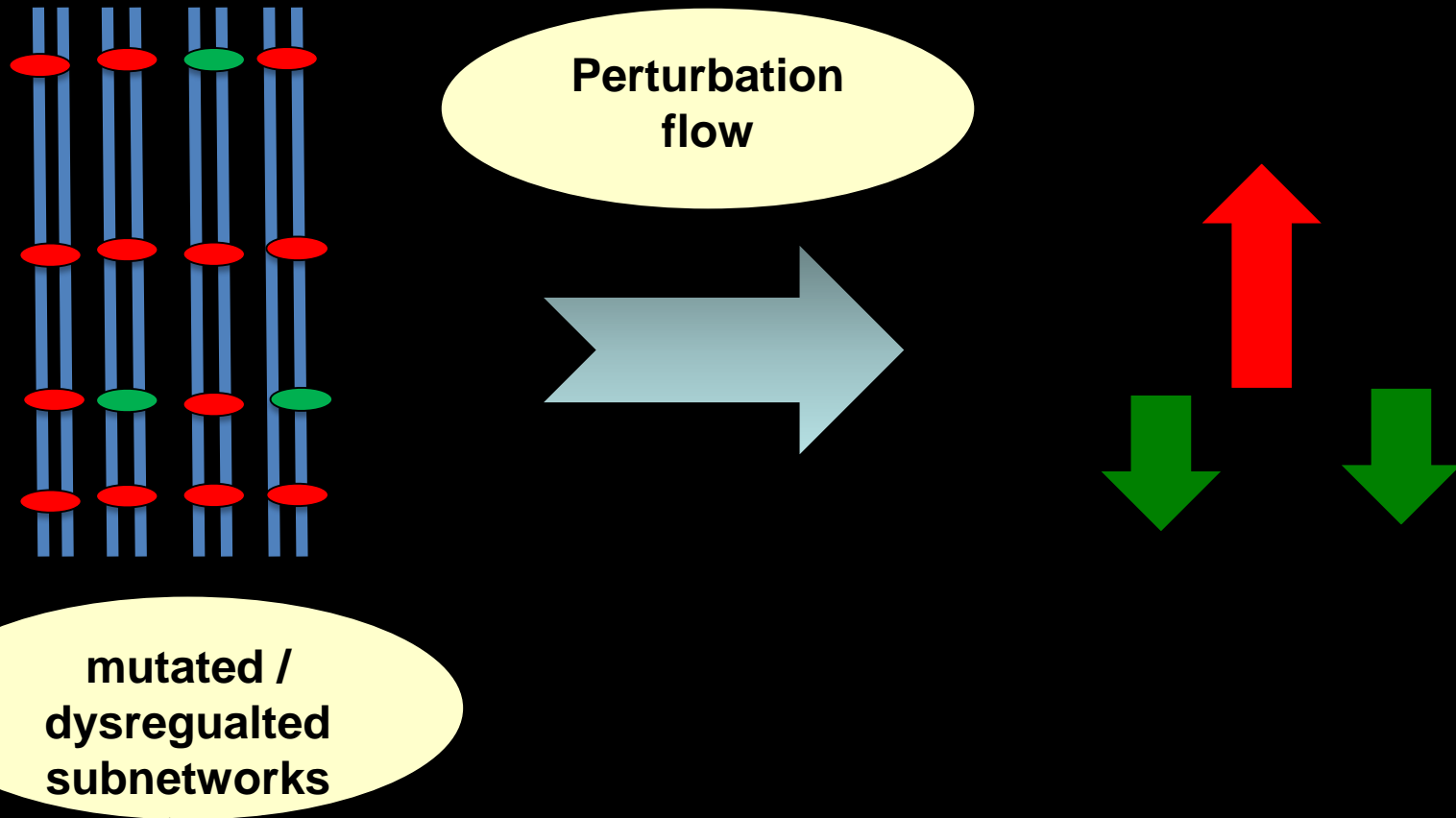
*Compared to Module Cover without ME*



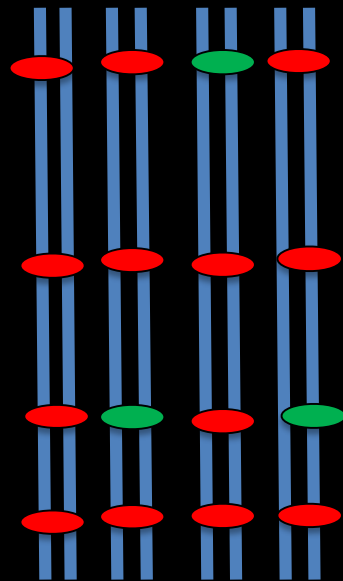
*Compared to HotNet2*



# Utilizing Networks for Understanding Genotype-Phenotype relations

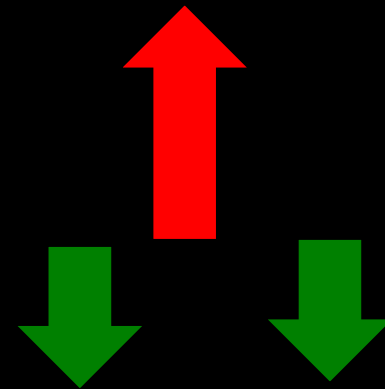
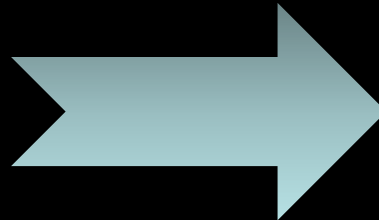


# Patient/phenotypic similarity networks



mutated /  
dysregulated  
subnetworks

Perturbation  
flow



phenotypic  
similarity  
networks

# Motivation – simultaneous utilization of multiple genotypic/causal and phenotypic data types

## Genotypic/ causal factors

mutations,

CNV

methylation,

Sex, age, environment ....

## Phenotypic properties

gene expression

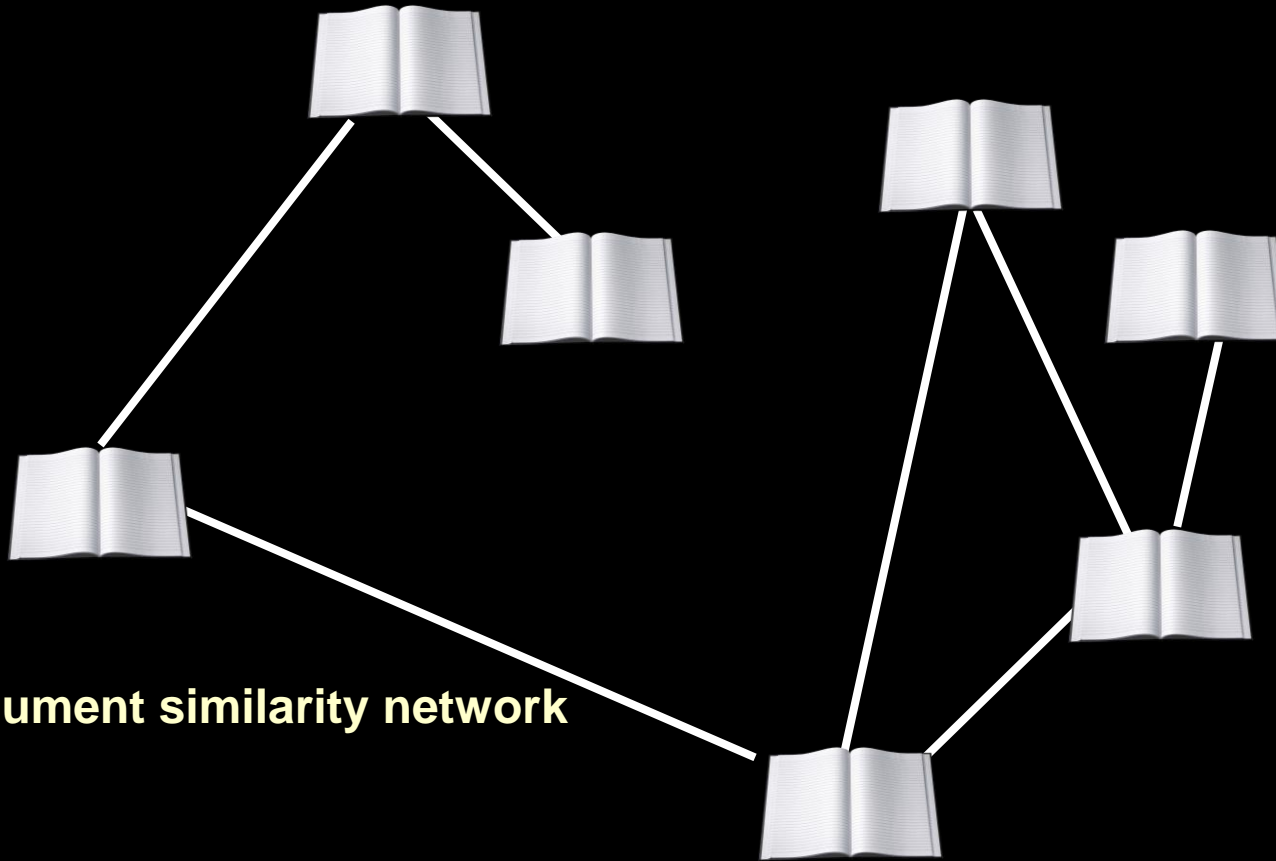
response to drugs

survival time

pathology features

**Idea:** Construct phenotype similarity graph and explain it connectivity using genotypic features

**Topic model:** Assuming that each document is a mixture of topics - identify topics and the words that define them



**Document similarity network**

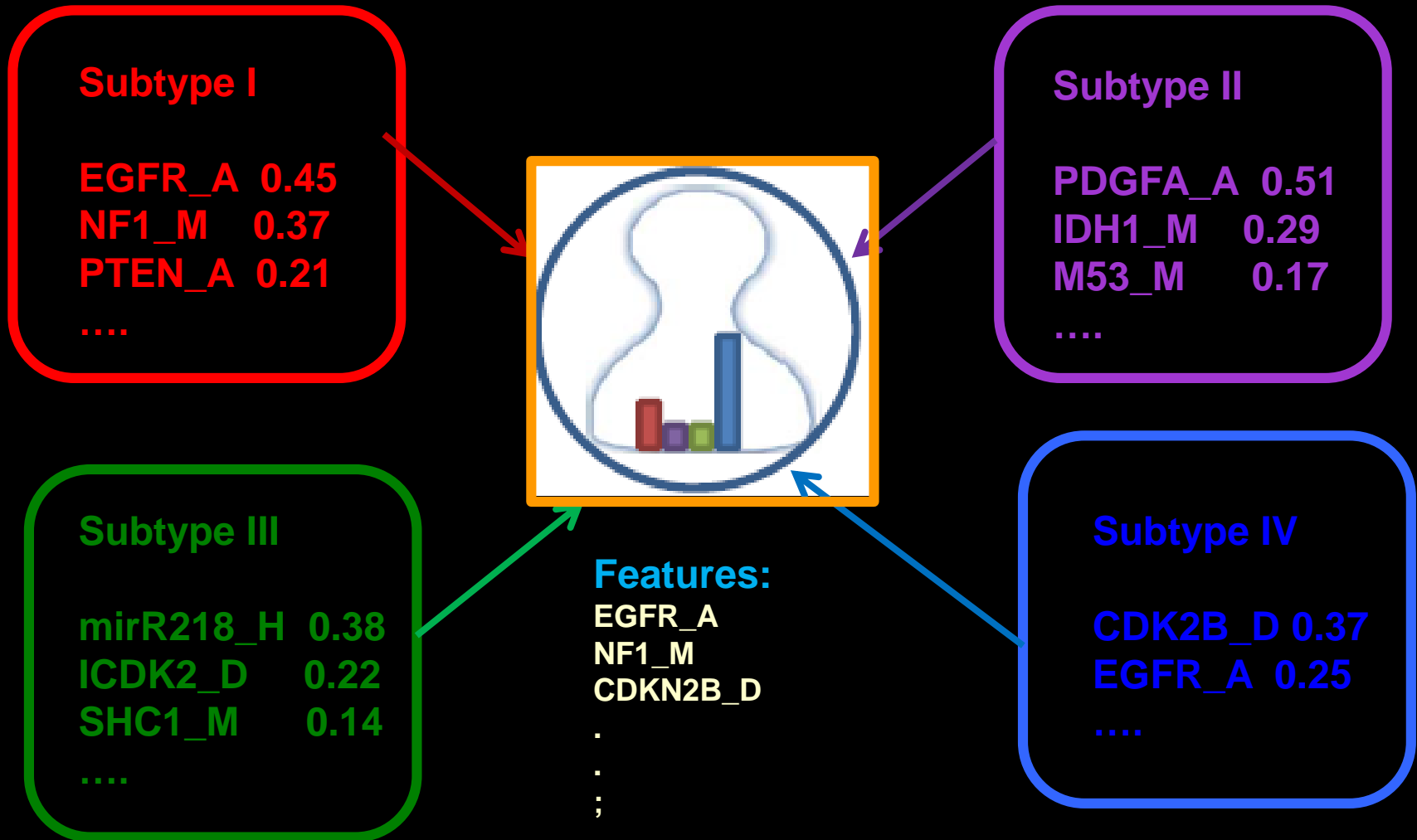
Chang J, Blei DM: Hierarchical Relational Models for Document Networks. Ann Appl Stat 2010, 4(1):124-150.



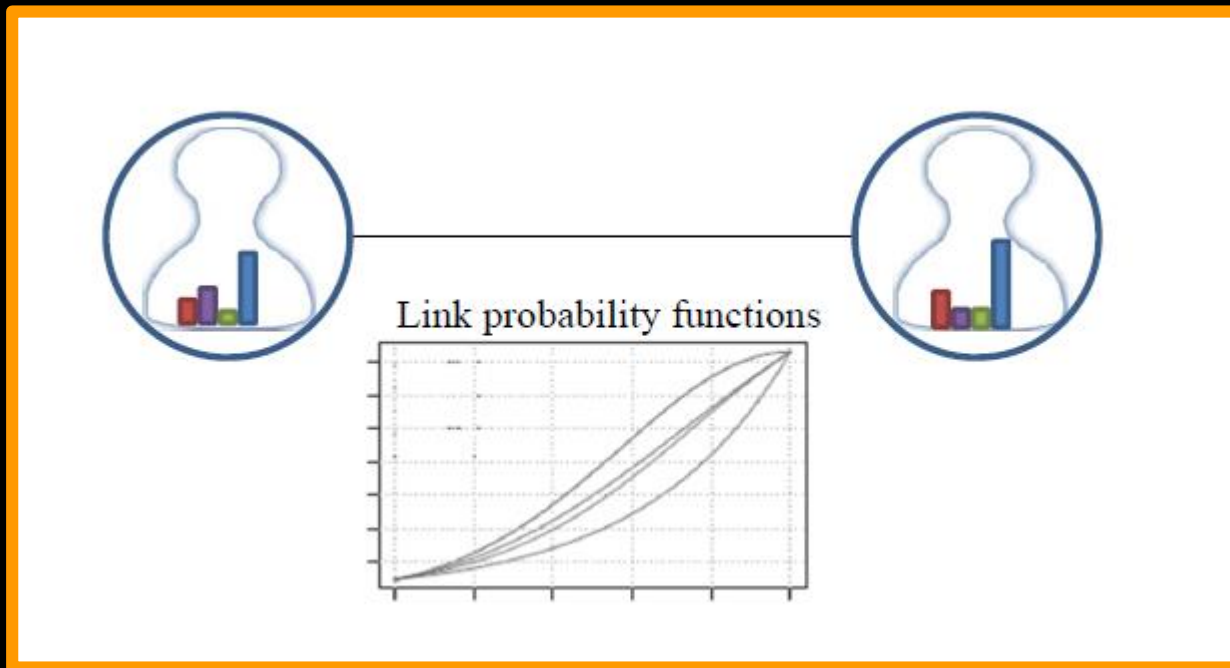
**Topics** – disease subtypes

**Words** – possible causes (mutations, CNV, miRNA)

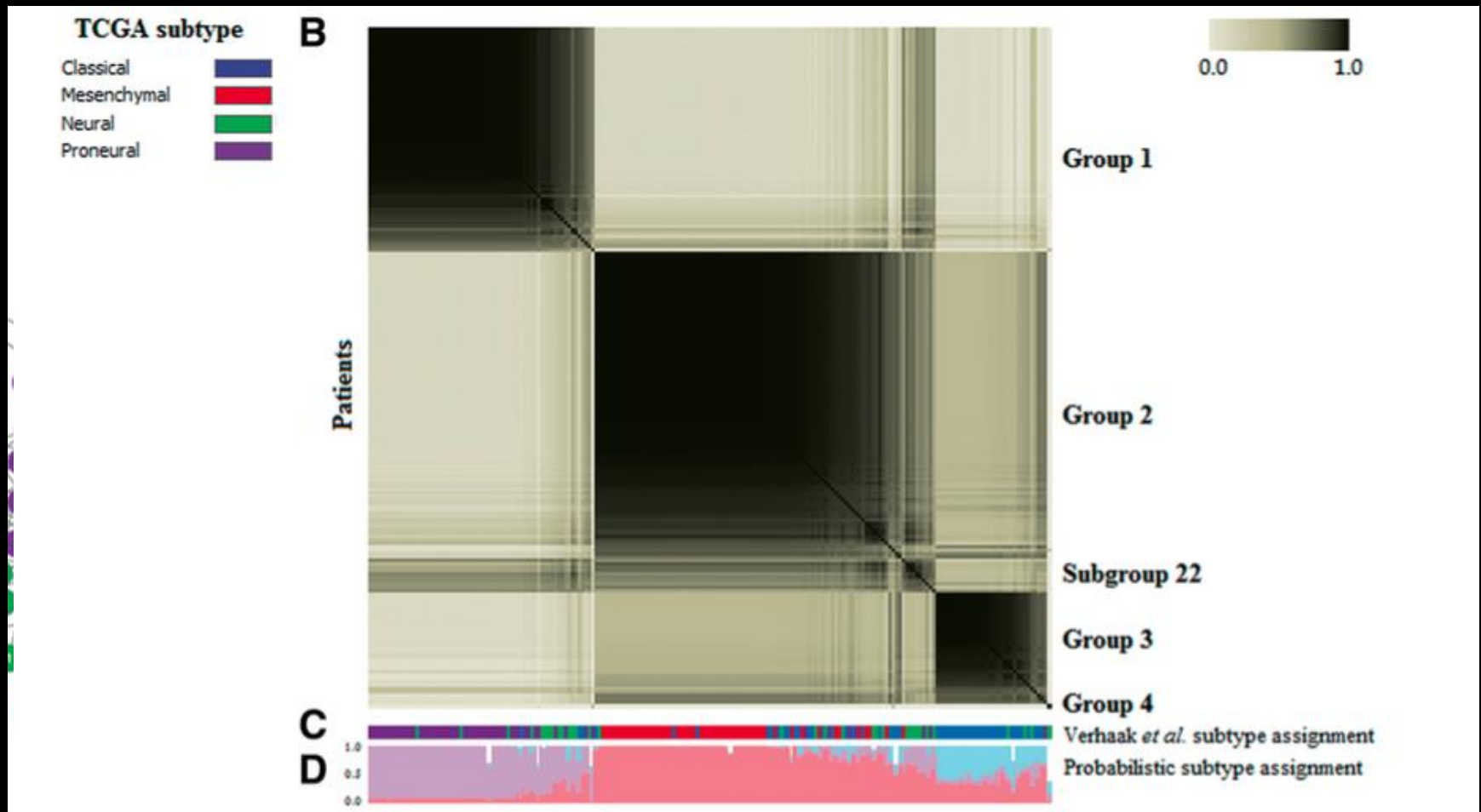
**Documents similarity**– phenotypic similarity (gene expression)



# Similar patients are to be explained by similar subtypes mixtures



# Co-occurrence of patients in the same subtype (based on 1000 topic models)



**Observation:** No separate Neural group

# Loss of Neural group is not surprising

Varhaak et al.  
Classification

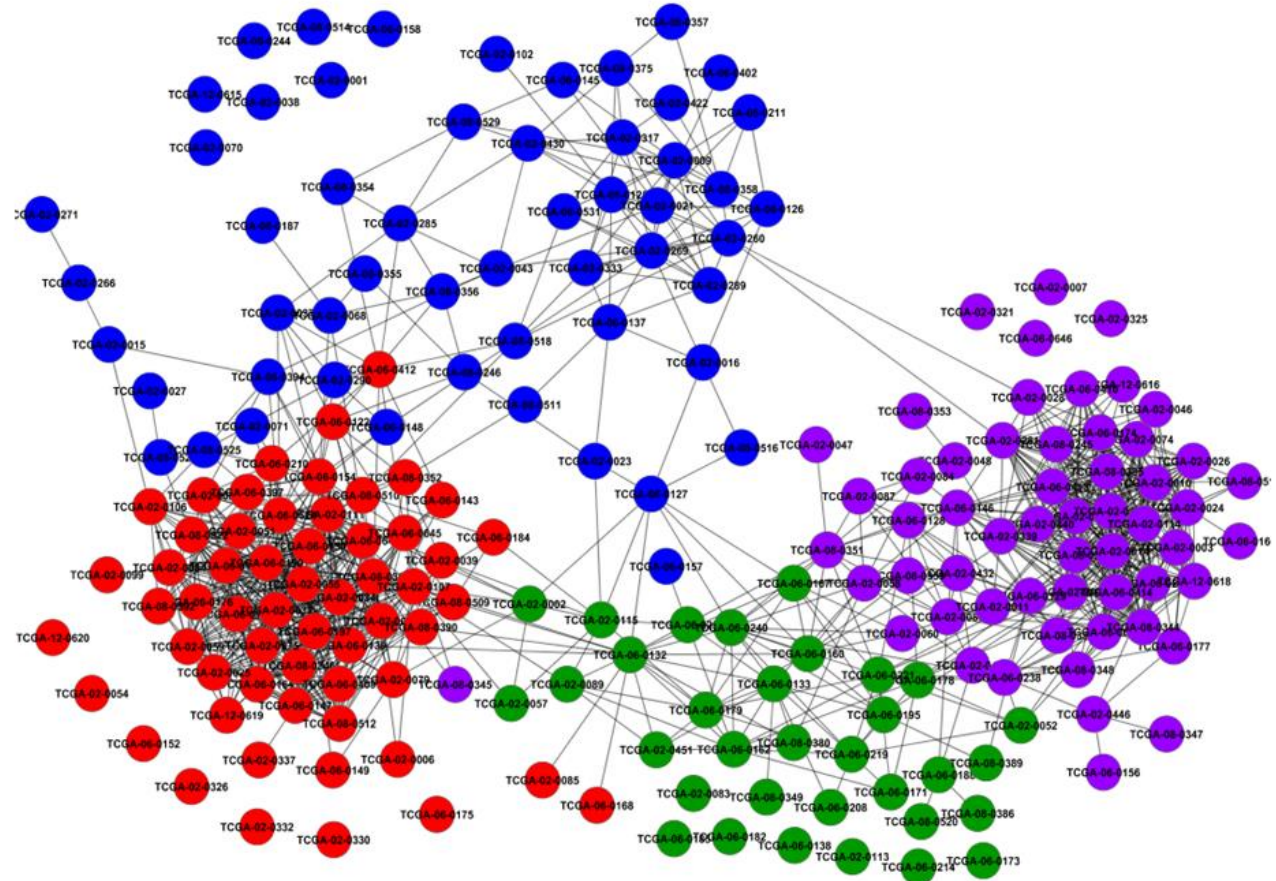
expression similarity network for GMB

 Mesenchymal

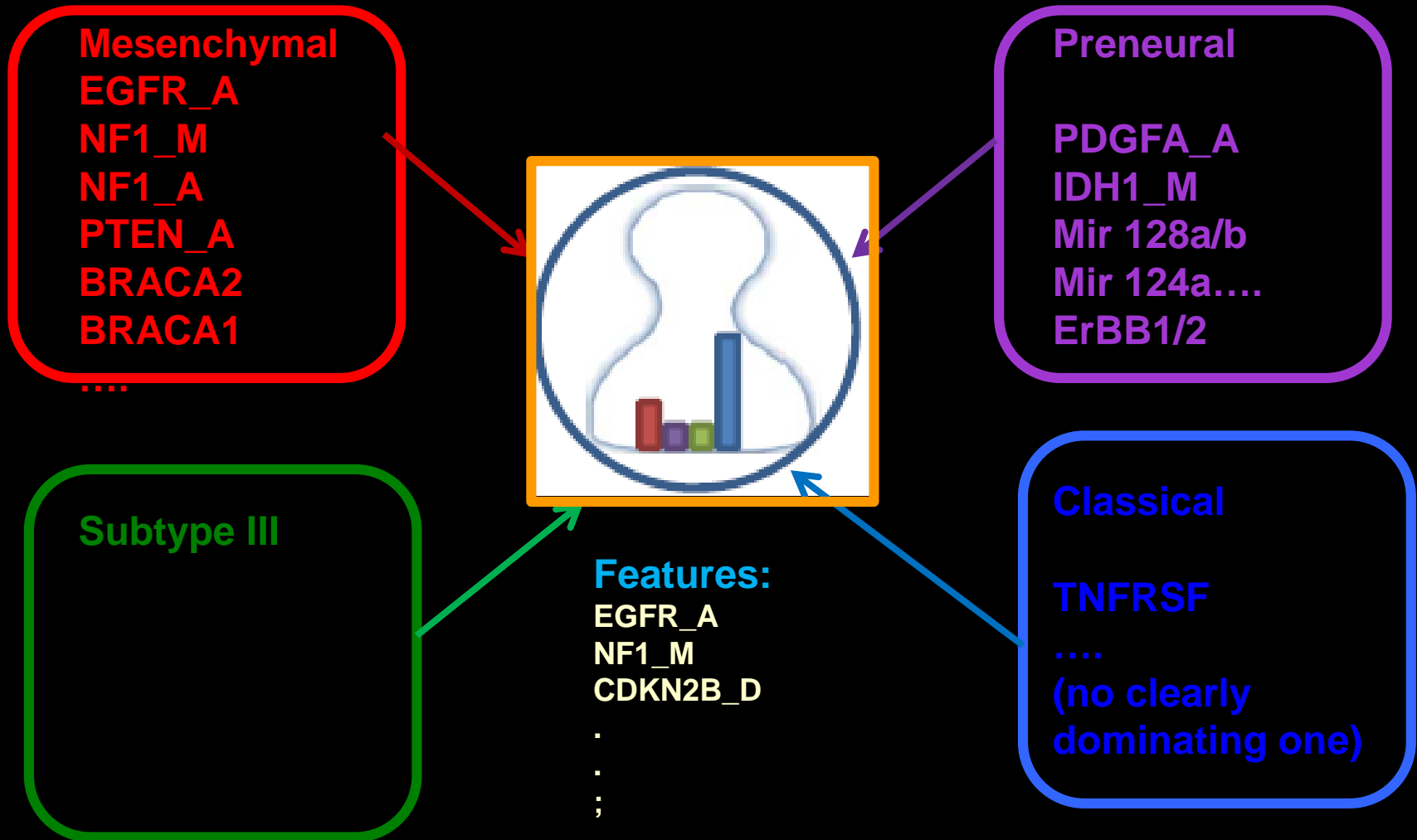
 Classical

 Proneural

 Neural



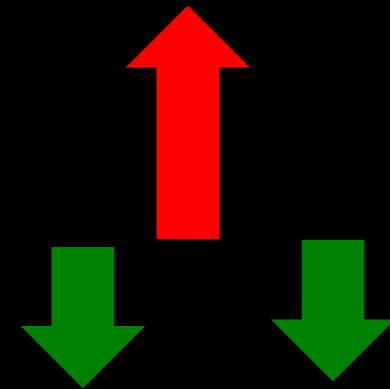
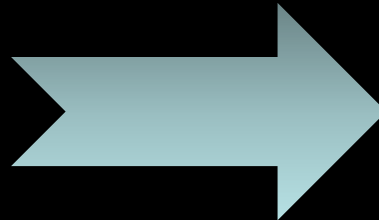
# Subtypes are defined by a distribution of words (mutations) – new patients can be easily classified



# Summary – these are complementing approaches

eQTL-net

Perturbation  
flow



phenotypic  
similarity  
networks

mutated /  
dysregulated  
subnetworks

Module-Cover

Topic Model





**Phung Dao**

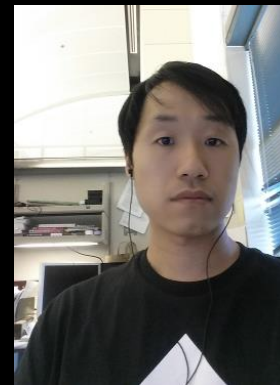
**Jan Hoinka**

**YooAh Kim**

**Damian Wojtowicz**

**Yijie Wang**

### Collaborators



**DongYeon Cho**  
(alumnae)

**Sanna Madam**  
Poolesville HS

**Brian Oliver**

**Hangnoh Lee**

Steve Russel, Cambridge  
Stefan Wuchty, Univ. Miami (not pictured)

# Perturbing a systems brings about valuable lessons

